

Shears, Beverly

73498

From: Devi, Sarvamangala
Sent: Friday, August 16, 2002 11:16 AM
To: Shears, Beverly
Subject: 09/921,157

Good morning Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 3 in application SN 09/921,157?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 42.28 Seconds

(without alignments)

3404.722 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMYRSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6554	98.2	1296	14	AA1980
2	6300.5	94.4	1287	16	AA1981
3	6220	93.2	1290	19	AA1982
4	6108	91.6	1288	18	AA1983
5	6108	91.6	1288	18	AA1984
6	2707.5	40.6	578	21	AA1985
7	1792.5	26.9	407	18	AA1986
8	1792.5	26.9	407	18	AA1987
9	1694.5	25.4	513	16	AA1988
10	1538.5	23.1	326	21	AA1989
11	1406	21.1	288	18	AA1990

12	1198	18.0	257	18	AAW55462	H. pylori ORF 07ap
13	1052	15.8	216	20	AAW89963	Antigen from clust
14	1029.5	15.4	224	21	AAW52506	Helicobacter pylori
15	861.5	12.9	176	18	AAW55346	H. pylori ORF 07ep
16	824	12.4	211	18	AAW55324	H. pylori ORF 07ap
17	824	12.4	211	18	AAW20755	H. pylori protein.
18	674	10.1	1484	22	AAW55686	H. pylori ORF 07ee
19	674	10.1	2902	22	AAW46351	H. pylori HPN165 p
20	648.5	9.7	2893	19	AAW98828	H. pylori GHPO 148
21	648.5	9.7	2893	19	AAW71556	Helicobacter poly
22	626	9.4	1382	19	AAW11001	H. pylori ORF 1lee
23	623	9.3	1413	18	AAW20725	H. pylori secreted
24	581	8.7	1974	19	AAW98391	H. pylori GHPO 57
25	569.5	8.5	786	18	AAW55464	H. pylori ORF 07ap
26	536	8.0	133	18	AAW55437	H. pylori ORF hp5p
27	509	7.6	560	18	AAW20409	H. pylori secreted
28	504	7.6	141	18	AAW20260	H. pylori protein.
29	504	7.6	141	18	AAW24621	H. pylori protein.
30	486	7.3	458	18	AAW55372	H. pylori ORF hp2e
31	486	7.3	458	18	AAW20964	H. pylori cytoplas
32	477.5	7.2	1213	18	AAW55735	H. pylori ORF 07ee
33	471	7.1	92	21	AAW52604	Helicobacter pylori
34	431	6.5	95	18	AAW55322	H. pylori ORF 07ap
35	431	6.5	95	18	AAW20754	H. pylori protein.
36	427	6.4	90	18	AAW20321	H. pylori protein.
37	421	6.3	93	18	AAW20362	H. pylori cytoplas
38	386	5.8	264	18	AAW20199	H. pylori cytoplas
39	320.5	4.8	1612	19	AAW65088	R. prowazekii S-la
40	289	4.3	1221	21	AAW01825	Haemophilus Influe
41	289	4.3	1227	21	AAW01824	Haemophilus Influe
42	288.5	4.3	1601	18	AAW30292	Non-typeable Haemo
43	287.5	4.3	1338	14	AAW41731	High molecular wei
44	287.5	4.3	1529	14	AAW41732	High molecular wei
45	284.5	4.3	1598	18	AAW30291	Non-typeable Haemo

ALIGNMENTS

RESULT 1
AAW41198
ID AAR41198 standard; Protein; 1296 AA.

XX AAR41198;

XX DT 17-MAR-1994 (first entry)

XX DE CT.

XX Cytotoxin; CT; H. pylori; precursor; vacuolation; cell death; hsp60;
KW heat shock protein; cytotoxin-associated immunodominant antigen; CAI;
KW hsp; type B; gastritis; peptic ulcer; eukaryote; gastric tumours.

OS Helicobacter pylori.

XX Key Location/Qualifiers

FT Peptide 1..33

FT Protein /note= "Signal peptide"

FT Protein 34..1296

FT Protein /note= "Mature Ct"

XX WO9318150-A.

XX PD 16-SEP-1993.

XX PF 02-MAR-1993; 93WO-EP00472.

XX PR 02-MAR-1992; 92IT-OFI0052.

XX PR 25-JAN-1993; 93WO-EP00158.

XX PA (BIOC-) BIOCINE SCLAVO SPA.

XX BUgnoli M, Covacci A, Macchia G, Rappuoli R, Telford J;

XX WPI; 1993-303464/38.
 DR N-PSDB; AAQ48732.
 XX
 PT Recombinant Helicobacter pylori protein and corresp. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and
 PT preventing type B gastritis, gastric ulcers and gastric tumours
 XX
 PS Claim 3; Fig 2; 83pp; English.
 XX
 CC This sequence represents the precursor form of cytotoxin, CT, of H.
 CC pylori. The DNA encoding this protein was isolated from two
 CC overlapping fragments corresponding to about 10 kb of the H. pylori
 CC genome. This protein has a molecular weight of 139.8 kD and serves as
 CC a precursor to a protein having an approximate weight of 100 kD and a
 CC cytotoxic activity. The amino acid positions 34-56 indicate a
 CC previously isolated N-terminal peptide. This sequence is preceded by
 CC 33 amino acids which resemble prokaryotic leader sequences, thus this
 CC sequence is likely to represent the N-terminal of the mature protein.
 CC A previously isolated cytotoxin of 87 kD may be produced by further
 CC processing or proteolytic degradation during purification. The mature
 CC cytotoxin causes vacuolation and death of a number of eukaryotic cell
 CC types. This protein, and others derived from H. pylori, esp.
 CC cytotoxin-associated immunodominant (CAI) antigen or a heat shock
 CC protein (hsp) (see also AAR41199-200), may be used to treat, prevent and
 CC diagnose H. pylori infection. H. pylori is the causative agent of type
 CC B gastritis, peptic ulcers and gastric tumours.
 XX
 SQ Sequence 1296 AA;

Query Match 98.2%; Score 6554; DB 14; Length 1296;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1277; Conservative 2; Mismatches 17; Indels 0; Gaps 0;

QY 1 MEIQOHRKINRPLVSLALVLCALVSLITPOQSHAAFFTVIIPALVGGTACTVAGTVSGL 60
 DB 1 melqqrhrknrlplvslalvslitpqqshaaftvtiipalvvggtatvagtvsyl 60

QY 61 LSWGLKQAEANKTPDKPKWRIQAGKGFNEFPNKEYDYRLSLSSKIDGWDGWNAAAR 120
 DB 61 lswglkqaeanktpdkpkwriqagkgfnefpnkeydyrlslsskidgwdwgnaar 120

QY 121 HYWKGGQONKLEVDMDKAVGTYTLGLRNFTGGDLVNMQKATLRLGQFNGNSFTSYKD 180
 DB 121 hywkggqonklevdmkavgttyslrlnftggdlvnmqkatlrlgqfngnsftsykd 180

181 SADRTTRVDFNAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGTTSDKNAEISLYD 240
 181 sadrttrvstlksqililqkstrvsgagrkasstvtlqasegitsdknaeisllyd 240

QY 241 GATNLASSSVKLMGNVNMGLQVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
 DB 241 gatnlasssvklmgvnmglqvaylapyystintskvtgevnfnhlvtvgdknaaqa 300

QY 301 GIANKTNTGTLDLWQAGLNIITAPPEGYKDKPNTPSQSGAKNDKNSAKNDKQESS 360
 DB 301 gikanntnigtldlwqaglniitappegykdkpntpsqsgakndknesakndkqess 360

QY 361 QNNSNTQVIPPNSAQTEVQPTQIDGPPFAGGKDTVVNIRINTNADGTLRVGGFKASL 420
 DB 361 qnnsntqvippnsaqtevpqtqidgppfaggkdtvvnirintnadgtlrvvgfkasl 420

QY 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANFEF 480
 DB 421 ttnaahlhigkggvnlngqasgrslivenltgntvdgplrvnnqvggyalagsanfef 480

QY 481 KAGTDTNKTATFNNDISLRFPVNLKVDATNFKGDITGNGGFNTLDFSCVTDKVNINK 540
 DB 481 kagtdtnktatfndislrfpvnlkvdatnfnkgditgnggfntldfsgvtdkvnink 540

QY 541 LITASTNVAVKFNINELIVKTNGISVGEYTHFSEDIQSQRINTVRLTGTSLFSGGV 600
 DB 541 litastnvavkfninelivktnngisvgeythfseidisqrintvrltgtslfsggv 600

DB 541 litastnvavkfninelivktnngisvgeythfseidisqrintvrltgtslfsggv 600
 QY 601 KFKGGEKLVIDEYFSPWNYFDARNIKNVEITNKLAFCPOCSPMGTSKLAFNNLTLQONA 660
 DB 601 kfkgeklvidefyspwnyfdarniknveitnklafcpgspwgtaklmfnnlitlqona 660
 QY 661 VMDYSQFSNLTIOGDFINNOGTFINYLVRGGKVATLSVGNAAAMFNNDIDSATFYKPLI 720
 DB 661 vmdysqfslntioqdfinnogtfinylvrggkvatlsvgnaaamfnndidsatfykpli 720

QY 721 KINSAQDLINKTEHVLKAKITIGYGVNSTGTNGISNVNLEOPKRLALYNNNRMDTCV 780
 DB 721 kinsaqdlinteavlakitigygvnstgtngisnvnleeqfkerfalylnnmnmtdtcv 780

QY 781 VRNTDDIKACGMAIGDOSVMNPNYKYLIGKAWKNIGISKTANGSKISVYVLGNSTPTE 840
 DB 781 vrntddikacgmaigdosvmnppnykyligkawnigisktangskisvvyvlgnstpte 840

QY 841 NGGNTTNPNTTNSARSAANALAQNAFPQPSATPNLVAIQNHDFGTIESVFELANRSK 900
 DB 841 nggnttnpntttnsarsannalaqnapfapqsatpnlvainqhdfgtiesvfelanrsk 900

QY 901 DIDTLYANSGAQRDLLOFTLLIDSHDAGYARKMIDATSANETIKOLNTATTTLNIIASLE 960
 DB 901 didtlyansgaqrdllloftllidshdagyarkmidatsaneitkolntatttlnniasle 960

QY 961 HKTSGLOTLISLNSAMILNSRLVNLRRHTNHIDSFARQLQAKDKQFASLESAAEVLVYQF 1020
 DB 961 hktsglqtlslisnamlinsrlvnlrrhtnhidsfarklqakdkqfasesaaevlyqf 1020

QY 1021 APKYEKPTNWANAIGTSTLNGSNASLYGTSAGVDAYLNGQVEAIVGGGSGSYGSSFNN 1080
 DB 1021 apkyekptnwanaigtstlngsnaalygtsagvdaylngvqaeivggfsgsygsfnn 1080

QY 1081 RANSLNGANNTFNGVYSRIFANOHEFDEAOGALGSDQSLNFKSALLODLNOSYHYLA 1140
 DB 1081 ranslnganntnfgvysrifanohefdeaqgalgsdqslnfksallqdlngsyhyla 1140

QY 1141 YSAATRASGYDFAFFRNALVLPKPSGVSVNHLGSTNFKNSTNOVALKNGSSSOHLFNA 1200
 DB 1141 ysaatrasgydfaffrnaivlpkpsgvsvnhlsgstnfknsntnovalkngsssqhlfna 1200

QY 1201 SANVEARYYGDTSYFYFNMAGVLQEFHAGVGSNNASLNTFKVNAARNPLNTHARVMGGE 1260
 DB 1201 sanvearyygdtsyfyfymagvlqefahvgsnnaaslnthfknvaarnplntharvmmgge 1260

QY 1261 LKLAKEVFLNLGVVYLHNLSINIGHFASNLGMRYSF 1296
 DB 1261 lklakevflnlgvvylhnlisnighfasnlgmrysf 1296

RESULT 2
 AAR79944
 ID AAR79944 standard; Protein; 1287 AA.
 XX
 AC AAR79944;
 XX
 DT 26-MAR-1996 (first entry)
 XX
 DE Helicobacter pylori vacuolating toxin.
 XX
 KW Vacuolating toxin; vaccine; immunisation; therapy; mutant;
 KW Infection; Helicobacter pylori.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9522988-A1.
 XX
 PD 31-AUG-1995.
 XX
 PF 23-FEB-1995; 95WO-US02219.
 XX
 PR 23-FEB-1994; 94US-0200232.

XX (UYVA-) UNIV VANDERBILT.
XX Blaser MJ, Cover TL;
XX WPI; 1995-311383/40.
XX N-PSDB; AAT04132.
XX Isolated DNA encoding Helicobacter pylori vacuolating toxin - useful
XX for immunisation against H. pylori infection
XX Claim 7; Page 43-47; 64pp; English.
XX The nucleic acid encoding the Helicobacter pylori vacuolating toxin
XX and a genetically altered mutant strain of H. pylori which contains
XX a foreign nucleic acid and does not express a functional vacuolating
XX toxin may be used to immunise a subject against H. pylori infection.
XX They may possibly also be used therapeutically.
SQ Sequence 1287 AA;
Query Match 94.4%; Score 6300.5; DB 16; Length 1287;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 1226; Conservative 29; Mismatches 32; Indels 9; Gaps 2;
Qy 1 MEIQTHRKINRPLVSLALVSLITPOOSHAAFTTIIIPAIVGGIATGTAVGTVSLG 60
Db 1 meiqqthrkinrplvslalvslitpqgshaaffttviiipaivvggiatgtavgtvsgl 60
Qy 61 LSWGLKQAEANKTPDKPKVMRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 lgwglkqaeeanktpdkpvriqagkgnfepnkeydlyrslsskiddgwdgnaat 120
Qy 121 HWYKGGQKLEVDKMDAVGTYTSLGRNFTGGDLVNMOKATLRLGQFNQNSFTSYKD 180
Db 121 hwykggqwnklevdmdkdvgtyslgrnftggdlvnmokatlrlgqfngnftsyykd 180
Qy 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLLQASGITSDKNAEISLYD 240
Db 181 sadrttrvdfnknslidnfnveinnrvsgagrkasstvtlltllqasgitssdknaeislvd 240
Qy 241 GATNLASSSVKLMGVNMGRLOYVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 gatnlasssvklmgvnmgrloyvgaylapsystintskvtgevnfnhlvtvgdknaaqa 300
Qy 301 GIANKKTHIGTLDLWQSLAGLNIAPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
Db 301 giasnktthigtldlwqslaglniappeggykdkpnntpsqsg-----akndkqess 360
Qy 361 QNNSNTQVINPPNSAOKTEVQPTQVIDGPFAGGKDPVNVINRINTNADGTRVGGFKASL 420
Db 361 qnnsntqvinnppnstqktvqptqvldgpfaggkdpvnnvinnrintnadgttrvvgfkasl 420
Qy 421 TTNAHLHTGKGVNLSNQAQSLRSLIVENLTGNTVDGRLRVNNQVGGYALAGSSANFEF 480
Db 421 ttnahlhtgkgvnlslnqaqslrslivenltgntvdgrlrvnnqvggyalagssanfef 480
Qy 481 KAGDTKNGTATFNNDISLGRFVNKLVDATNFAGKIDTNGGFFNTLDFSGVTDKVNINK 540
Db 481 kagvtdkngtatfnndisigrfvlvklvdahtnfnagkidtnggffntldfsgvtknvink 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEIDIGSOSRINTVRLTGTSLFSGGV 600
Db 541 litastnvavkfninelivkngisvgeythfseidigsosrintvrltgtslfsggv 600
Qy 601 KFKGSEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPGWTSKLMFNNTLQONA 660
Db 601 kfkgselkvidefyfswpnyfdarniknveitnklafpgqspgwtsklmfnntlqona 660
Qy 661 VMDYSQFSLNLTIGQDFINNOGIIYVIRGKGVATLSVGNAAAMFNNDISATGFKPLI 720
Db 661 vmdysqfslnltigqdfinnogiiyvirgkgvatlsvgnaaamfnndisatgfkpli 720
Qy 721 KINSAODLIKNTHEHVLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 721 kinsadlikntehvllkakiigygnvstgtngisnvnleeqkerlalynnnrmdtcv 780
Qy 781 VRNTDDIKACGMAIGQDSVMVNNPDNYKYELGRKAWKIGISKTAGSKISVYVYLGSTPTE 840
Db 781 vrntddikacgmaigqdsvmvnnpdnykyeligrakwignishtagkskisyvylgnspte 840
Qy 841 NGCNTNTNLPTNTSARSANALAOAPPAQPSATPNLVAINOHDFCTIESVFLANRSK 900
Db 841 ngcntntnlptntsnarsanalaopaapqpsatpnlvainohdfctiesvflanrsk 900
Qy 901 DIDTLVANSAGCROLLQTLILLIDSHDAGYARKMIDATSAEITKQKLTATTTLNNTIASLE 960
Db 901 didtlvansagcqdllqtlillidshdagyartmidatsaneitkqintattnlniasle 960
Qy 961 HKTSGIQTLSLSNAMTLNLSRLVNLRRHIDSFRAKRLQALDKQKFALESAAEVLVYQF 1020
Db 961 hktssiqtlslslsnamlslvnlrrhidsfraqalqldkqrfaslesaaevlyqf 1020
Qy 1021 APKYKPTNVWANAIGTSLNNGSNASLYCTSGAGVDAYLNGOVEAIVGGFGSYGYSFNN 1080
Db 1021 apkykptnvwanaiagtslnngsnaslyctsgavdaylngoveaivggfsgysfnn 1080
Qy 1081 RANSLNSGANNTNFGVYSRIFANQHEFDFEAQCALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Db 1081 ranslsganntnfgvysrifanqhefdfeaqcalgsgdsslnfkallrdlnqsynyla 1140
Qy 1141 YSAATRASGYDPAFFRNALVLKPSVGVSYNHLGSTNFKSNSTNQVALKNGSSQHLFNA 1200
Db 1141 ysaatrasgydaffrnalvlkpsvgvsynhlgstnfksnstnkvalngssqhlfna 1200
Qy 1201 SANVEARYYGDPSYFYMNAGVLOEFAHYGSSNNAASINTFKVNAARNPLNTHARVMGGE 1260
Db 1201 sanvearyygdpsyfymnagvlqefahygssnnaasintfkvnaarnplntharvmgge 1260
Qy 1261 LKLAKVFFNLGVVY 1296
Db 1261 lklakevfnlgvvvylhnlisnighfasnlgmrysf 1296
RESULT 3
AAW98269
ID AAW98269 standard; Protein; 1290 AA.
XX
AC AAW98269;
XX
DT 31-MAR-1999 (first entry)
DE
DE H. pylori GHPO 374 protein.
XX
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
XX WO9843478-Al.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
XX 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
DB

DR N-PSDB; AAX13988.

XX New isolated Helicobacter polynucleotides - used to develop products for the diagnosis, prevention and treatment of Helicobacter infections and gastrointestinal diseases

XX Claim 8; Page 273-278; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPO protein of the invention. The polypeptides can be used for preventing or treating Helicobacter infections, and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

XX Sequence 1290 AA;

Query Match 93.2%; Score 6220; DB 19; Length 1290;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1207; Conservative 38; Mismatches 45; Indels 6; Gaps 2;

QY 1 MEIOOTHRKINRPLVSLALVGLYSITPQOSHAAFFTTVIIPAIVGGIATGATVGTSGSL 60
DB 1 melqgthrklnrpvlslalvglvsltpqgshaaffttvliipaivvggiatgaavgtvsgl 60

QY 61 LSWGLQAEAEANKTPDKPKVRIQAKGKGFNEFPNKEYDLYRSLSSKIDGWDGWAAR 120
DB 1 lgwllqaeaeanktpdkpkvriqakgkgnfepnkeydlyrslsskldgwdgwnaat 120

QY 121 HYWKGGQONKLEVDMDKAVGTYTSLGRNFTGGDLVNMOKALRLGQFNGNSFTSYKD 180
DB 121 hywkkggqonklevdmkdavgtvslgrnftggdlvnmokalrlgqfngnftszykd 180

QY 181 SADTRTRVDFNAKNISIDNFEIINRVGSGAGKASSVLTQLQASEGITSKPKAEISLYD 240
DB 181 sadtrtrvdfnknislidnfeimrvvgsgagkassvltqlqasegitsskpkaeislyd 240

QY 241 GATNLASSSVKLMGNVMGRLOVGYALAPSYTINTSKVTGEVNFHNLVGVGDKNAQA 300
DB 241 gatnlaasssvklmgvnmgrlovgyalapsytintskvtgevnfnhnlvvgdknaqa 300

QY 301 GIIANKRTNIGTLDLWQAGLNIITAPEGGYKDKPNTPSOSGAKNDKNSAKNDKBQSS 360
DB 301 giiasnkthigtldlwsaglniitappegykdkpdkps-----nttqnannnqgnsa 355

QY 361 QNNSNTQVINPNSAQKTEVQPTVIDGPPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
DB 356 qnnsntqvinpnsaqkteiqptvidgpfaggkdtvvnnirintnadgtirkvgykaskl 415

QY 421 TTNAHLHIGKGVNLSNQASGRSLIVENLTGNTVDGPLRVNQQVGYALAGSSANFEF 480
DB 416 ttnaahlhigkgvlnsqasgrslivenltnvtdgplrvnqqvgyalagssanfef 475

QY 481 KAGTDTKNGTATFNNDISLGRFVNKLVDYAHFANFKGIDTGNCGNTLDFSGVTDKVNINK 540
DB 476 kagtdtkngtatfnndislgrfvnklvdyahfannfkgidtgnngntldfsgvtdkvnik 535

QY 541 LITASTNVAVKFNFINELIVKTINGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 600
DB 536 litastnvavknfinelivktingisvgeythfseidigsqsrintvrltgtslfsggv 595

QY 601 KFKGGEKLVIDEFYSPNYPDARNIKNVEITNKLAFGPQSGPWGTSKLMFNNTLQGNA 660
DB 596 kfkgeklvidefyyspnyfarniknveitrkfasstpenpwtgsklmfnntlqgna 655

QY 661 VMDYSQESNLTQDGFINNQCTINLYRGKGVATLSVGNAAAMFNNDIDSATGFYKPLI 720
DB 656 vmdysqesnltdgfdinnqctinlyrgkgvatlsvgnaaamfnndidsatgfykplm 715

QY 721 KINSAQDLIKNTEHVLLKAKIIGYCNVSTGTGINSNVNLEQFKERLALYNNNNRMDTCV 780
DB 721 kinsaodliknteHVLLKAKIIGYCNVSTGTGINSNVNLEQFKERLALYNNNNRMDTCV 780

DB 716 kinsaodliknkehvllkakliigynvslgtnsisnvnllleqfkerlalylnnnnrmldcv 775

QY 781 VRNTDDIKACGMAIQDQSMVNNPNYKYLICKAWKNIGISKITANGSKISVYVLGNSTPTE 840
DB 776 vrntddikacgtalngqsmvnnpnnykyligkawnigistangskislvyvlgnstpte 835

QY 841 NGGNTNLPTNTTNSARSNANALAQNAFPAQPSATPNLVAINQHDGFGTIESVFELANRSK 900
DB 836 kggntnlptnttnsarsannalagnapaqpsatpnlvainqhdfgtiesvfelelansk 895

QY 901 DIDTLYANSQAQRDLQTLTLLIDSHDAGYARKMIDATYSAEITTKQLNTATTTLLNNIASLE 960
DB 896 didtlyansqagrdlltllidshdagyarqmdntstgkqlnaatllnniasle 955

QY 961 HKTSGLQTLTSLSNAMIINSLRVLNLSRRHTNHIDSFARKLQALQDKQKFALESAAEVLQYF 1020
DB 956 hktsslqtlslsnamiinlsrlvlnlsrrhtnnidfaqralkdkqkfaslesaeelyqf 1015

QY 1021 APKYEKPTNWWANAIGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSGYSSFN 1080
DB 1016 apkyekptnwwanalggtslnngsnaslygttsagvdaylngvealvggfgsgyssfnn 1075

QY 1081 RANSLNCANNNTNFGVSRIFANOHEPDEFAQAGALGSDOSLSLNFKSLALLOLNOSYHYLA 1140
DB 1076 qanslncannntnfgvsrifanqhefdefaagalgdsoslslnfnksalrlngsynyla 1135

QY 1141 YSAATRASGYDFAFFRNALVLPKPSVGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1200
DB 1136 ysaatrasgydaffrnalvlpkpsvgvsnhlgstnfksns-nqvalkngsssqhlfna 1194

QY 1201 SANVEARYYIGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
DB 1195 sanvearyyigdtstyfyminagvloefahvgssnnaaslntharvmmgge 1254

QY 1261 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1296
DB 1255 lklakevflnlgvylhnlisnighfnsmrysf 1290

RESULT 4
AAW5547
ID AAW55547 standard; Protein; 1288 AA.
AC AAW55547;
XX XX
DT 24-JUN-1998 (first entry)
XX XX
DE H. pylori ORF 14ee41924_2458267_c2_93 secreted protein.
XX XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteric; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
XX XX
PN W09737044-A1.
XX XX
PD 09-OCT-1997.
XX XX
PF 27-MAR-1997; 97WO-US05223.
XX XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX XX
PA (ASTR) ASTRA AB.
XX XX
PI Alm RA, Smith D;
XX XX
DR WPI; 1997-503122/46.
DR N-PSDB; AAV24956.

XX Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Disclosure; Page 750-753; 1145pp; English.
XX
XX This sequence is a H. pylori secreted protein. The protein may be used
CC in a vaccine to prevent or treat H. pylori infection or to identify
CC H. pylori polypeptide binding compounds, useful as potential H. pylori
CC life cycle activators or inhibitors. The DNA and probes derived from it
CC may be used for the identification of H. pylori in a sample and the
CC diagnosis of H. pylori infection. Nucleic acid sequences complementary
CC to the DNA act as antisense sequences and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can be
CC used in immunoassays to evaluate the abundance and distribution of
CC H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 1288 AA;

Query Match 91.6%; Score 6108; DB 18; Length 1288;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1181; Conservative 53; Mismatches 54; Indels 8; Gaps 2;

Qy 1 MEIQTHRKINRPLVSLALVGLVSLITFCQSHAAFTTVPVPAIVGGTATGATVTSGL 60
Db 1 meiqthrkirplvslalagalsaiqbqshaftftviiipaivggatgtavgtvsgl 60
Qy 61 LSWGLQAEANKTPDKPKWRIQAGKGFNPKVEYDLVRLSLSSKIDGGMDGNAR 120
Db 61 lswglkqaeanktpdkpkwriqagkgnfepkneydlvrlslsskldggmdgwnar 120
Qy 121 HYWKGQONKLEVDMDKAVGTYTSLGRNFTGGDLVNMQKATLRLGQFNQNSFTSYKD 180
Db 121 hywkggqwnklevdmdkavgtyklsgrnftggdlvnmqkatlrlgqfngnsftsykd 180
Qy 181 SADRTTRVDFNAKNISIDNFVNNRVSGAGRKASSVTLTQASEGITSQKNAEISLYD 240
Db 181 sadrttrvfnaknisidnfvninnrvsgagrkassvtltqasegitssknaeislyd 240
Qy 241 GATLASSSVKLMGNVWMLQYGVAYLAPSYSTINTSKVTGEVNFHLTVGDKNAQA 300
Db 241 gatllansvklngnvmwmlrlyqygaylapsystintskvgevdhltvvgdqaqa 300
Qy 301 GIANKTNIGTDLQWAGLNIAPPPEGKYDKPNNTPSQSGAKNDKNAESAKNDQESS 360
Db 301 giiasknthigtldlwqaglniapppegkydkpnntpsqsgakndknaesakndqess 360
Qy 361 QNNSNTQVINPPNSAQKTEVQPTQIDGPFAGGKDTVVNINRINTNADGTRVGFKASL 420
Db 361 -nnsntevinnpntqketeptqidgpfaggkdtvvnnfnhntkadgtlkvggfksl 413
Qy 421 TTNAALHLHGKGVNLSNOAGRSILVNLGNINVDGPLRVNNOVGGVALAGSSANFEF 480
Db 421 ttnaalhlhgkgvnlsnoagrsilvnlgninvdgplrvnnovggvalagssanfef 473
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAAHTANFKGIDTNGGFFNTLDFSGVTDKVNINK 540
Db 474 kagvtdkngtatfnndisigrfvlkvdahntanfkidtgngffntldfsgvtdkvnink 533
Qy 541 LITASTNVAVKFNINELIVKTNGISVGEYTHFSDEIGSQSRINTVRLTGTGRSLFSGGV 600
Db 541 litastnvavkfninellivktnngisvgeythfsedigsqsrintvrltgrtsrfsfsgv 593

Db 534 litastnvavkfninellivktnngisvgeythfsedigsqsrintvrltgrtsrfsfsgv 593
Qy 601 KFKGGEKLVIDDFYVSPWNPYDARNIKVVEITNKLAFGPQGSQWCTSKLWFWNLTLGONA 660
Db 594 kfksgelivindfyyspwnyfdarnvknvelcrkfasstpenpwtgsklmfnntltgqna 653
Qy 661 VMDYSQFSLNLTIOGDFINNOGTINTLVRRGGKVATLSVGNAAAMFNNDISATGYKPLI 720
Db 654 vmdysqfslntltgqdfinnqgtintlvrrgkvatlnvgnaaamfnndidsatgyfkpli 713
Qy 721 KINSQADLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEFKRLALYNNNNMDTCV 780
Db 714 kinsaqdlikntehvllkakiliygnvstgtngisnvnleefkerlalynnnnmdtcv 773
Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKNIGISTANGSKTSVYVLGNSTPTE 840
Db 774 vrntddikacgmaignqsmvnnpdnykyligkawnigisktangsktsvvyignstpce 833
Qy 841 NGGNTNLTPTNTSARNANALAQNAFPAQSPATPNLVAIINOHDFGTIESYFELANRSK 900
Db 834 nggntnltptntnnaahsanyalvknafah-satpnlvainqhdfgtiesvfelanrsk 892
Qy 901 DIDTLIYANGSAGGRLDQLTLLIDSHDAGYARKMIDATSAEITKOLNTATTTLNIAISLE 960
Db 893 didtlythsgagrdldlqtlidshdagyardmtdntstgeitkqlnaatdalmnvase 952
Qy 961 HKTSGLQTLISLNSAMILNSRLNLSRRHTNHDSEAKRLQALKDOKFASLESAAEVLQYF 1020
Db 953 hkqsglqtlislnamilsrlnlsrrkthnhsfaqlralkqgefasesaaevlyqf 1012
Qy 1021 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSSFN 1080
Db 1013 apkyekptnvwanalggaslnsgsnaslygtsgvdaflngvnealvggfygssfn 1072
Qy 1081 RANSLNSCANNTNFGYSRIFANQHEFDEFAOGALGDSOSSLNFKSALLQDLNQSHYLA 1140
Db 1073 qanslnsgannanfgyrsfrfnghefdeafgdsqsslnfstlqldngshyila 1132
Qy 1141 YSAATRASYGDFAFERNALVLKPSGVSYNHLGTSNFKSNSTNOVALKNGSSOHLFNA 1200
Db 1133 ysatarasygydfafernalvlkpsgvsyhlgstnfksnsgvalkngassghlfn 1192
Qy 1201 SANVEARYYGDTSYFYMNAGVLQBFHAGVHVSNNAAASLNTFKYNAARNPLNTHARVMMGE 1260
Db 1193 nanvearyyygdsfyfahagvlqbfhagvhsndvaslntfkinaarsplstaryammge 1252
Qy 1261 LKLAKEVELNLGVVVLHNLISNIGHFASNLGMRYSF 1296
Db 1253 lqlakevfinlgvvyhlhnlisnashfasnlgmrysf 1288

RESULT 5
AAW55685
ID AAW55685 standard; Protein; 1288 AA.
XX
AC AAW55685;
XX
DT 07-JUL-1998 (first entry)
XX
DE H. pylori ORF 07ee11402_2458267_c3_108 cell envelope OMP.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX

XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW bait polypeptide; gastric ulcer; antibacterial.
XX Helicobacter pylori.
OS Helicobacter pylori.
XX WO2000066722-A1.
XX 09-NOV-2000.
XX 14-APR-2000; 2000WO-IB00603.
XX 30-APR-1999; 99EP-0401066.
XX (HYBR-) HYBRIGENICS SA.
XX Legrain P, Selig L, Rain J;
PI WPI; 2000-687535/67.
XX N-PSDB; AAC97321.
XX A two-hybrid system for identifying compounds useful in the treatment
of e.g. gastric ulcers comprises producing a collection of recombinant
cell clones -
PT
XX
XX Example 5; Page 202-204; 267pp; English.
XX The present sequence is a bait polypeptide used in a Helicobacter
XX pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX Sequence 578 AA;
SQ
Query Match 40.6%; Score 2707.5; DB 21; Length 578;
Best Local Similarity 90.1%; Pred. No. 7.8e-149;
Matches 525; Conservative 27; Mismatches 26; Indels 5; Gaps 1;
QY 182 ADRTTRVDNNAKISIDNFEINNRVSGAGRKASSTVLTLOASEGITSKDKNAEISLYDG 241
DB 1 adrttrvdfnakilidnfeinnrvsgagrkasstvtlqasegitskksnaeislydg 60
242 ATLNASSSVKLMGNVWGRLOYGAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQAG 301
DB 61 atlnlasvsklmgnvwgrloygaylapsystintskvtgevnfnhnlvtvgdhnnaaqag 120
QY 302 ITANKTNTGTLDMOSAGLNIAPPEGYKDKPNWTPSQSGAKNDKESAKNDKQESSQ 361
DB 121 iasankthigtldlwsaglniappegykdkpdkps-----nttgnannngnqnsaq 175
QY 362 NNSNTQVINPPNSAQTEVQPTQVIDPPAGGKDTVWNNRINTNADGTRVGGFKASLT 421
DB 176 nnsntqvinppnsaqteiqptqvidppfagkgdvtvnnidrtntnadtgtikvgykaslt 235
QY 422 TNAHLHICKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNOVGYSALAGSSANFEFK 481
DB 236 tnaahlhikgginlsnqasgrtlivenltgnitvdgplrvnnqvgysalagssanfek 295
QY 482 AGTDTKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTNGGFGENTLDFSGVTDKNINKL 541
DB 296 agtdtkngtatfnddislgrfvnlkvdahtanfkidtgngfntldfsgvtgkvtnkl 355
QY 542 ITASTNVAVKNFNNELIVKTNIGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGVK 601
DB 356 itastnvavknfnlnelivktnigisvgeythfseidigsqsrintvrltgttrslfsggvk 415

QY 602 FKGEKLVIDEFYISPNYFDARNIKNVEITNKLAFCPOGSPMGTSKLMFNLTIGONAV 661
DB 416 fksgeklviddefyyspwnyfdarniknveitrkfaststpenwgtkslmfnltlignav 475
QY 662 MDYSQFSNLTIOGDFINNOGTINYLVRGKGVATLSYGNAAAAMFNNDIDSATGYKPLIK 721
DB 476 mdysqfsnltioqdfinnogtlnylvrggvatlnvgnaaamfnsnvdsatgyfqlmk 535
QY 722 INSAQDLIKNTHVLKAKIIGYGNVSTGTNGISNVNLEEQFK 764
DB 536 insaqdliknkehvlkakligygnvslgtnsisnvnlieqfk 578
RESULT 7
AAW55290
ID AAW55290 standard; Protein; 407 AA.
XX
AC AAW55290;
XX
DT 15-JUN-1998 (first entry)
XX
DE H. pylori ORF 02ep30607orf19 protein.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX Helicobacter pylori.
XX WO9737044-A1.
XX 09-OCT-1997.
XX 27-MAR-1997; 97WO-US05223.
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX (ASTR) ASTRA AB.
XX Alm RA, Smith D;
XX WPI; 1997-503122/46.
DR N-PSDB; AAV24699.
XX Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX Claim 14; Page 524-525; 1145pp; English.
XX This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds. The
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 35679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in *E. coli*
 XX hosts.
 SQ Sequence 407 AA;

Query Match 26.9%; Score 1792.5; DB 18; Length 407;
 Best Local Similarity 89.2%; Pred. No. 4.9e-96;
 Matches 346; Conservative 17; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEIOOHRKINRPLVSLALVGLVSIPTQOSSHAAFFTTVIIPAIVGGTATCTAVGTYSGL 60
 DB 17 meiqdthrklnrpvlsvlagallisaipqeshaaffttviiipaivggatgtavgtvsgl 76
 QY 61 LSWGLKQAEAEANKTPDKPKVWRIOAGKGFNEFPNKEYDYLSLSKIDGWDGNAAR 120
 DB 77 lswglkqaeaeanktpdkpkvwrliqagkgfnefpnkeydyllskidgwdgwnaar 136
 QY 121 HYWYKGGQONKLEVDMDKDAVGYTTLGSLRNFTGGDLVNMQKATLRLGQFNNGSFTSYKD 180
 DB 137 hywvkggqwnklevdmdkdvgytklslrnfthgdlvnmqkatlrlgqfngnsftsykd 196

QY 181 SADRTTRVDFNAKNSIDNFEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
 DB 197 sadrttrvfnaknsidnfeinrnrvsgagrkasstvtltqasegitssknaeislyd 256
 QY 241 GATNLASSSVKLMGNVMMGRLOVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
 DB 257 gatlnlsssvklnvnmgrlqvgyalapsystintskvgevdfnhlvtvgdqnqaqa 316
 QY 301 GIANKKTNIGTLDLWOSAGLNIIAPPEGGYKDKPNNTPSOSGAKNDKQESS 360
 DB 317 gilasnkthigtldwlsaglniilappeggykdkpnstsgtkndkksqn----- 370

QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDG 388
 DB 371 -nnsntevinnppnntqkteteptkslmg 397

RESULT 8
 AAW20652
 ID AAW20652 standard; protein; 407 AA.

XX AC AAW20652;

DT 14-JUL-1997 (first entry)

XX H. pylori derived protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

OS Helicobacter pylori.

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Melligaard BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT67905.

PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*

PT infection, and to detect *Helicobacter*
 XX Disclosure; Page 1074-1075; 1481pp; English.

XX The present sequence is a *H. pylori* derived protein of unspecified
 CC function, no further details are given in the specification.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds,
 CC useful as potential *H. pylori* life cycle activators or inhibitors.
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely *H. pylori* antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
 CC production, e.g. in *E. coli* hosts.

XX Sequence 407 AA;

Query Match 26.9%; Score 1792.5; DB 18; Length 407;
 Best Local Similarity 89.2%; Pred. No. 4.9e-96;
 Matches 346; Conservative 17; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEIOOHRKINRPLVSLALVGLVSIPTQOSSHAAFFTTVIIPAIVGGTATCTAVGTYSGL 60
 DB 17 meiqdthrklnrpvlsvlagallisaipqeshaaffttviiipaivggatgtavgtvsgl 76

QY 61 LSWGLKQAEAEANKTPDKPKVWRIOAGKGFNEFPNKEYDYLSLSKIDGWDGNAAR 120
 DB 77 lswglkqaeaeanktpdkpkvwrliqagkgfnefpnkeydyllskidgwdgwnaar 136

QY 121 HYWYKGGQONKLEVDMDKDAVGYTTLGSLRNFTGGDLVNMQKATLRLGQFNNGSFTSYKD 180
 DB 137 hywvkggqwnklevdmdkdvgytklslrnfthgdlvnmqkatlrlgqfngnsftsykd 196

QY 181 SADRTTRVDFNAKNSIDNFEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
 DB 197 sadrttrvfnaknsidnfeinrnrvsgagrkasstvtltqasegitssknaeislyd 256

QY 241 GATNLASSSVKLMGNVMMGRLOVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
 DB 257 gatlnlsssvklnvnmgrlqvgyalapsystintskvgevdfnhlvtvgdqnqaqa 316

QY 301 GIANKKTNIGTLDLWOSAGLNIIAPPEGGYKDKPNNTPSOSGAKNDKQESS 360
 DB 317 gilasnkthigtldwlsaglniilappeggykdkpnstsgtkndkksqn----- 370

QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDG 388

DB 371 -nnsntevinnppnntqkteteptkslmg 397

RESULT 9
 AAR79945
 ID AAR79945 standard; protein; 513 AA.

XX AC AAR79945;

XX 26-MAR-1996 (first entry)

XX Helicobacter pylori antigenic protein.

XX Vacuolating toxin; vaccine; immunisation; therapy; mutant;

XX Infection; *Helicobacter pylori*.

XX Helicobacter pylori.

XX WO9522988-A1.

```
PD 31-AUG-1995.
XX
PF 23-FEB-1995; 95WO-US02219.
XX
PR 23-FEB-1994; 94US-0200232.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Blaser MJ, Cover TL;
XX
XX WPI: 1995-311383/40.
DR N-PSDB; AAT04133.
XX
XX Isolated DNA encoding Helicobacter pylori vacuolating toxin - useful
PT for immunisation against H. pylori infection
XX
XX Claim 19; Page 50-52; 64pp; English.
XX
XX Nucleic acid encoding the Helicobacter pylori vacuolating toxin
CC and a genetically altered mutant strain of H. pylori which contains
CC a foreign nucleic acid and does not express a functional vacuolating
CC toxin may be used to immunise a subject against H. pylori infection.
CC They may possibly also be used therapeutically. This antigenic
CC protein may also be used for immunisation purposes.
XX
XX Sequence 513 AA;
SQ

Query Match 25.4%; Score 1694.5; DB 16; Length 513;
Best Local Similarity 63.3%; Pred. No. 3.2e-90;
Matches 343; Conservative 52; Mismatches 84; Indels 63; Gaps 7;

QY 311 GTLDLWQASAGLTIAPPEGYKDKPNTPSQGAKNDKSNKAKNDQESSQNNSNTQVIN 370
Db 1 gvtldlwqasaglsitppegysktdkpsqsknd----- 37

QY 371 PPSNAQTEVQPTQVIDGPFAGGKDTVNNINRINTNADGTVRGFKASLTNNAAHLHG 430
Db 38 ----Lqkteiqvtldgpfaggkdtvnnifhlntkadgtlkaggfkaslttnaaahlhg 93

QY 431 KGGVNLNQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEFKAGDTKNGT 490
Db 94 eggvnlnsqasgrslivenltgnitvegtlrvnnqvggaaavagssanfefkagadtnnat 153

QY 491 ATFNNDISLGRVNLKVAHT-----ANFKGIDT--GNGGFNT- 526
Db 154 atfnndihlgkavnrvdahtayfngniylgkstnlrvnghshahfkndatksdnglnts 213

QY 527 -LDFSGVTDKVNINKLITASTWAVKFNINELIVKTNGISVGEYTHFSEDIGSQSRINT 585
Db 214 aldfsgvtdkvninkltsatvnnknfdikelvvttrvqsfgytiffgenigdksrigv 273

QY 586 VRLGTGTRSLFSGGVKFGKEGLVIDEYFYSPPNYFDARNIKNVEITNKLAFGQSGSPWG 645
Db 274 vrlgtgyspaysgvtfksgklvldelnyhapwnyfdarnvtdveinkrillfgapgniaq 333

QY 646 TSKLMFNLLTQGNVMDYSQFSNLTIOGDFINNOGTINYLVRGKVAATLSVGNAAAMMF 705
Db 334 ktglmfnnltlnsasmgykdlditigbftcnngtmtalfvqgdgrvatlnaghqasmif 393

QY 706 NNDIDSATGFYKPLKINSADOLIKNTEHVLKAKIIGYGNVSTGTNG-----ISNVN 758
Db 394 naldsatsgfykplkinnaglnktnkehlvlgkgrnidnlyv--gvqgasyndisasntn 451

QY 759 LEEQKERLALYNNNRMDTCVVR--NTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKN 816
Db 452 lleeqkerlalylnnnrmdicvrkgntddikacgmaignqsmvnnpndykyledkawkn 511

QY 817 IG 818
Db 512 tg 513
```

```
RESULT 10
AAB52500
ID AAB52500 standard; Protein; 326 AA.
XX
AC AAB52500;
XX
XX 23-FEB-2001 (first entry)
XX
XX Helicobacter pylori bait polypeptide #18.
DE
DE Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW bait polypeptide; gastric ulcer; antibacterial.
XX
XX Helicobacter pylori.
OS
XX WO2000066722-A1.
PN
PD 09-NOV-2000.
XX
XX 14-APR-2000; 2000WO-IB00603.
XX
XX 30-APR-1999; 99EP-0401066.
XX
XX (HYBR-) HYBRIGENICS SA.
PA
XX Legrain P, Selig L, Rain J;
XX
XX WPI: 2000-687535/67.
DR N-PSDB; AAC97246.
XX
XX A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
XX Example 5; Page 101; 267pp; English.
XX
XX The present sequence is a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
XX Sequence 326 AA;

Query Match 23.1%; Score 1538.5; DB 21; Length 326;
Best Local Similarity 90.3%; Pred. No. 1.9e-81;
Matches 299; Conservative 15; Mismatches 12; Indels 5; Gaps 1;

QY 207 VGSAGRKASSTVLTQLQASEGITSDKNAEISLYDGTATLNASSVKLMGNWMLRQYVG 266
Db 1 vsgagrkasstvtlqasegitssknaeislydgatlnasvklmgvnmgrlqyvg 60

QY 267 AYLAPESTINTSKVTGEVNFNHLTVGDKNAQAQGIANKKTWIGTLDLWQASAGLNIAP 326
Db 61 aylapstintskvtgevnfnhltvgdhnhaqagliasnktthigtldlwqsaglniap 120

QY 327 PEGGYKDKPNTPSQGAKNDKSNKAKNDQESSQNNSNTQVINPNSAQKTEVQPTQVI 386
Db 121 peggykdkpdkdps-----nttqpnannhqqsagqnsntqvinnpsaqkteiqptqvi 175

QY 387 DGPFFAGGKDTVNNINRINTNADGTVRGFKASLTNNAAHLHGKGVNLNQASGRSLI 446
Db 176 dgpfaggkdtvnnidrintnadtgtikvggkaslttnaaahlhgkgnlnsqasgrtll 235

QY 447 VENTLGNITVDGPLRVNNQVGGYALAGSSANFEFKAGTGTKNGTATFNNDISLGRVNLK 506
Db 176 vnltnitvtdgplrvnnqvggylagssanfefkagtgatfnndisigrvnlk 506
```

Db 236 ventlgnitvdgplrvnnqvggyalagssanfefkagtdkngtatfndndislrgrfvnlk 295
 Qy 507 VDAHTANFKGIDTNGGFGNTLDFSGVTDKVN 537
 |||||
 Db 296 vdahtanfkgidtgngfgntldfsgvtgkvn 326
 |||||
 RESULT 11
 AAW27720
 ID AAW27720 standard; Protein: 288 AA.
 XX
 AC AAW27720;
 XX
 XX 11-MAY-1998 (first entry)
 DT
 DE H. pylori VacA protein autotransporter region.
 XX
 XX VacA gene; autotransporter; diagnostic; therapy;
 KW Gram-negative bacteria; surface presented polypeptide.
 PS Helicobacter pylori.
 PH Key Location/Qualifiers
 FT Protein 1..288
 FT /note= "partial protein sequence"
 XX
 XX WO9735022-A1.
 XX
 XX 25-SEP-1997.
 PD
 XX 15-MAR-1996; 96WO-EP01130.
 PF
 XX 15-MAR-1996; 96WO-EP01130.
 PR
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA
 XX Jose J, Maurer J, Meyer TF;
 PI
 XX WPI; 1997-480227/44.
 DR N-PSDB; AAT88157.
 XX
 XX Presentation of peptide(s) on surface of Gram-negative bacteria -
 PT via transformation with vector encoding signal peptide, presented
 PT peptide and transporter domain of auto-transporter, producing
 PT peptide libraries for epitope mapping
 XX
 XX Claim 8; Fig 24; 84pp; German.
 PS
 CC This sequence represents an autotransporter membrane integration
 CC region from the H. pylori vacA gene. This region is involved in a novel
 CC method which allows the presentation of stable fusion polypeptides on the
 CC surface of Gram-negative bacteria which can be released into the
 CC surrounding media. The method can be used to produce a variegated
 CC population of surface-presented polypeptides, so that bacteria expressing
 CC polypeptides with particular properties can be identified and
 CC simultaneously selected, e.g. for epitope mapping or selection of ligands
 CC with the highest affinity for antibodies, major histocompatibility
 CC complex (MHC) molecules or other components of the immune system.
 CC Selected polypeptides can be used diagnostically, e.g. to screen sera or
 CC antibody banks, and polypeptide expressing cells may be used as live
 CC vaccines. They may be used therapeutically, e.g. when the polypeptide is
 CC an antibody, to remove or concentrate pollutants, inactivate toxins,
 CC prepare and process food, prepare washing compositions and label cells.
 CC Selected bacteria can be stored, reproduced and replicated on a large
 CC scale as individual clones.
 XX
 XX Sequence 288 AA;
 SQ
 Query Match 21.1%; Score 1406; DB 18; Length 288;
 Best Local Similarity 94.1%; Pred. No. 7.6e-74;
 Matches 271; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

Qy 1009 SLESAAEVLQFAPKYEKPTNVWANAIGGTSLNNGSNASLYGTSGVDAYLNGOVEAIVG 1068
 |||||
 Db 1 slesaaevlyqfapkyekptnnwanaiggtcslnggghnaslygtsgvdaylungevealvg 60
 |||||
 Qy 1069 GFSGSYGYSSENNRANSLNSGANNNTFCVYSRIFANQHEFOFEAGALGSDOSSLNFKSAL 1128
 |||||
 Db 61 gfsgsygysfnsqanslmsgannncfgyvysrifanqhefdeaqga lgsdqsslnfkksal 120
 |||||
 Qy 1129 LQDLNQSYHYLAYSAATRASVGYDFAFFRNALVLKPSVGVSYNHLGSTNFKSNSTNOVAL 1188
 |||||
 Db 121 lrdlnqsynylaysaatrasvgydfaffrnalvlkpsvgyvsnhlgstnfnksnqkval 180
 |||||
 Qy 1189 KNGSSQHLFNASANVEARYYYGDTSYFYMNAGVLQBEFAHVGSNNAASLNTFKYNAARNP 1248
 |||||
 Db 181 kngassqhlfnasanvearyyygdtvfyymnagvlqefanfngssnavslntfkvnatrn 240
 |||||
 Qy 1249 LNTHARVMMGELKLAKEVFLNGLGVVYLHNLISNIGHFASNLGMRYSF 1296
 |||||
 Db 241 lntharvmmgdelklakevflnlgfvyhlhlnisnighfaasnlgmrysf 288
 |||||
 RESULT 12
 AAW55462
 ID AAW55462 standard; Protein: 257 AA.
 XX
 AC AAW55462;
 XX
 XX 24-JUN-1998 (first entry)
 DT
 XX
 DE H. pylori ORF 07apl1015_23938312_f3_2 cell envelope OMP.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 XX WO9737044-A1.
 PN
 XX 09-OCT-1997.
 PD
 XX
 XX 27-MAR-1997; 97WO-US05223.
 PF
 XX 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 XX (ASTR) ASTRA AB.
 PA
 XX Alm RA, Smith D;
 PI
 XX WPI; 1997-503122/46.
 DR N-PSDB; AAV24871.
 XX
 XX Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptides(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 XX
 PS Disclosure; Page 667-668; 1145pp; English.
 XX
 CC This sequence is a H. pylori cell envelope outer membrane protein
 CC (OMP) having a terminal phe residue.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of


```
XX The present sequence is a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX Sequence 224 AA;
SQ
Query Match 15.4%; Score 1029.5; DB 21; Length 224;
Best Local Similarity 88.2%; Pred. No. 3.5e-52;
Matches 202; Conservative 10; Mismatches 12; Indels 5; Gaps 1;
165 LRLQFGNGNSFTSYKDSADRTTRVDFAKNKISIDNFVEINRVGSGAGRKASSTVLTQA 224
Db 1 lrlqfgngnstsykdsadrttrvdfaknildnfleimrvsgagrkasstvlqla 60
Qy 225 SEGITSDKNAISLYDGA TLNLASSVKLMGNVWGR LQYVYGAYLAPSYSYNTSKVTGE 284
Db 61 segitssknaeislydgatlnlasnsvklmgvnmgrlqyvgaylapsystintskvtge 120
Qy 285 VNFNHLTVGDKNAQAAGIIANKNTNIGTLDLWQSAGLNIITAPPEGYKDKPNNTPSQSGA 344
Db 121 vnfnhltvgdhnaaqaagilasnkthigtldlwqsaglnilappgegykdkpkdps---- 176
Qy 345 KNDKESAKNDKQSSQNNSTQVINPPNSAQKTEVQPTQVIDGPFAGG 393
Db 177 -ntqnnannqnsaqnnsntqvinppnsaqkteiqptqvidgpfagg 224
RESULT 15
AAW55346
ID AAW55346 standard; Protein; 176 AA.
XX
AC AAW55346;
XX
XX 17-JUN-1998 (first entry)
XX
XX H. pylori ORF 07ep30818orf4 protein.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacteria; life cycle; activator;
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
XX
XX WO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
XX 02-APR-1996; 96US-0758731.
XX 25-OCT-1996; 96US-0736905.
XX 28-OCT-1996; 96US-0738859.
XX
XX (ASTR ) ASTRA AB.
XX
XX Alm RA, Smith D;
XX
XX WPI; 1997-503122/46.
XX N-PSDB; AAV24755.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
```

```
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claim 14; Page 570-571; 1145pp; English.
XX
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunosays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
XX hosts.
XX Sequence 176 AA;
SQ
Query Match 12.9%; Score 861.5; DB 18; Length 176;
Best Local Similarity 93.8%; Pred. No. 1.4e-42;
Matches 166; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
Qy 701 AAMFNNDDIDSATGFGYKPLIKINSAQDLIKNTEHVLKAKTIIGYGVSTGTNGISNVNLE 760
Db 1 aamfnnddidsatgfykplikinsaqdliknte hvlkakiigynvstgtngisnvnle 60
Qy 761 EQFKERIALYNNNNRMDTCVVRNTDDIKACGMAIGDQSMVNNPDNYKYLGKAWKNIGIS 820
Db 61 eqfkerialynnnnrmtdtcvvrntddikacgmalignqsmvnnpdnykylgkawrnlgis 120
Qy 821 KTANGSKISVYILGNSTPTTENGNTNLPTNTTNSARNALAQNAFPQPSATPN 877
Db 121 ktangskisvvyilgnstptengntnlptnttnnhsanyalvknafah-nttpn 176
```

Search completed: August 16, 2002, 11:30:35
Job time: 191 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 18.86 Seconds
(without alignments)
1678.450 Million cell updates/sec

Title: US-09-921-157-3
Perfect score: 6671
Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	6671	100.0	1296	3	US-08-470-260-3
2	6671	100.0	1296	3	US-08-471-491-3
3	6671	100.0	1296	4	US-08-466-662-3
4	6303.5	94.5	1287	1	US-08-200-232-2
5	6303.5	94.5	1287	5	PCT-US95-02219-2
6	6303.5	94.5	1287	5	PCT-US95-02219A-2
7	1695.5	25.4	513	1	US-08-200-232-4
8	1695.5	25.4	513	5	PCT-US95-02219-4
9	1695.5	25.4	513	5	PCT-US95-02219A-4
10	320.5	4.8	1612	1	US-08-169-927-2
11	292	4.4	1529	2	US-08-728-470-10
12	292	4.4	1529	4	US-08-719-641-10
13	287.5	4.3	1338	2	US-08-728-470-9
14	287.5	4.3	1338	4	US-08-719-641-9
15	287.5	4.3	1338	2	US-08-719-641-9
16	283	4.2	1541	4	US-08-296-791-3
17	283	4.2	1541	5	PCT-US95-10661A-3
18	276	4.1	1545	4	US-08-296-791-4
19	276	4.1	1545	5	PCT-US95-10661A-4
20	274	4.1	1600	2	US-08-617-697-10
21	269	4.0	1536	1	US-08-038-682-2
22	269	4.0	1536	1	US-08-302-832-2
23	269	4.0	1536	2	US-08-530-198-2
24	269	4.0	1536	2	US-08-469-880-2
25	269	4.0	1536	2	US-08-728-470-2
26	269	4.0	1536	2	US-08-617-697-2
27	269	4.0	1536	4	US-08-719-641-2

28	268	4.0	2048	4	US-09-268-347-48	Sequence 48, Appli
29	265.5	4.0	1702	4	US-08-296-791-5	Sequence 5, Appli
30	265.5	4.0	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
31	258.5	3.9	2314	4	US-09-268-347-49	Sequence 49, Appli
32	248	3.7	1477	1	US-08-038-682-4	Sequence 4, Appli
33	248	3.7	1477	1	US-08-302-832-4	Sequence 4, Appli
34	248	3.7	1477	2	US-08-530-198-4	Sequence 4, Appli
35	248	3.7	1477	2	US-08-469-880-4	Sequence 4, Appli
36	248	3.7	1477	2	US-08-728-470-4	Sequence 4, Appli
37	248	3.7	1477	2	US-08-617-697-4	Sequence 4, Appli
38	248	3.7	1477	4	US-08-719-641-4	Sequence 4, Appli
39	244	3.7	2353	4	US-09-377-155-33	Sequence 33, Appli
40	244	3.7	2353	4	US-08-913-942-4	Sequence 4, Appli
41	244	3.7	2353	4	US-09-669-974-33	Sequence 33, Appli
42	243	3.6	1912	1	US-08-409-995-4	Sequence 4, Appli
43	243	3.6	1912	3	US-08-685-467-4	Sequence 4, Appli
44	242	3.6	2354	4	US-09-268-347-47	Sequence 47, Appli
45	238	3.6	1394	4	US-08-296-791-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-470-260-3
; Sequence 3, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-260-3

Query Match 100.0%; Score 6671; DB 3; Length 1296;

	Best Local Similarity	100.0%	Pred. No. 0:	Matches 1296:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
Qy	1	MEIQTHRKNRPLVSLALV	60	SHAAFTFTTIIPIAVGGIATGTA	60							
Db	1	MEIQTHRKNRPLVSLALV	60	SHAAFTFTTIIPIAVGGIATGTA	60							
Qy	61	LSWGLKQAEANKTPDKPKV	120	WVRQAQGFNEFPNKEVDYLRSLSSKIDG	120							
Db	61	LSWGLKQAEANKTPDKPKV	120	WVRQAQGFNEFPNKEVDYLRSLSSKIDG	120							
Qy	121	HYWKGGQONKLEVDMDK	180	AVGTYTILSGLRNFTGGDLVNMOKATLRLQ	180							
Db	121	HYWKGGQONKLEVDMDK	180	AVGTYTILSGLRNFTGGDLVNMOKATLRLQ	180							
Qy	181	SADRTTRVDFNAKNISID	240	NEFVEINNRVSGAGRKASSTVLTQLQASEGIT	240							
Db	181	SADRTTRVDFNAKNISID	240	NEFVEINNRVSGAGRKASSTVLTQLQASEGIT	240							
Qy	241	GATLNLAASSVKLMGNV	300	MGRLOYGVAYLAPSYSTINTSKYTGVEYFNH	300							
Db	241	GATLNLAASSVKLMGNV	300	MGRLOYGVAYLAPSYSTINTSKYTGVEYFNH	300							
Qy	301	GIIANKTNICTLDLQW	360	SAGLNIITAPPEGGYKDKPNNTPSOGAKNDK	360							
Db	301	GIIANKTNICTLDLQW	360	SAGLNIITAPPEGGYKDKPNNTPSOGAKNDK	360							
Qy	361	QNNSTQVINPPNSAQKTE	420	VOPTVIDPPAGGKDTVYINRINTNADGTRV	420							
Db	361	QNNSTQVINPPNSAQKTE	420	VOPTVIDPPAGGKDTVYINRINTNADGTRV	420							
Qy	421	TTNAHLHIGKGVNLS	480	NOAGRSILVENLTGNTIDGPLRVNNOVG	480							
Db	421	TTNAHLHIGKGVNLS	480	NOAGRSILVENLTGNTIDGPLRVNNOVG	480							
Qy	481	KAGDTKNGTATFNNDIS	540	LGRFVNLKVDAHTANFKGIDTNGGNFTL	540							
Db	481	KAGDTKNGTATFNNDIS	540	LGRFVNLKVDAHTANFKGIDTNGGNFTL	540							
Qy	541	LITASTNVAVKFNINELI	600	VTNGISVGEYTHFSEDIGSQSRINTVRL	600							
Db	541	LITASTNVAVKFNINELI	600	VTNGISVGEYTHFSEDIGSQSRINTVRL	600							
Qy	601	KFKGGEKLIVDEFYYP	660	WNYPDARNIKNVEITNKLAFPGQSPWGT	660							
Db	601	KFKGGEKLIVDEFYYP	660	WNYPDARNIKNVEITNKLAFPGQSPWGT	660							
Qy	661	VMDYSQFSNLTIOGDF	720	INNOCTINFLVGGKVATLSVGNAAAMFN	720							
Db	661	VMDYSQFSNLTIOGDF	720	INNOCTINFLVGGKVATLSVGNAAAMFN	720							
Qy	721	KINSAQDLIKNTEHVLL	780	KAKIIGYGNVSTGNGISNVNLEEQFK	780							
Db	721	KINSAQDLIKNTEHVLL	780	KAKIIGYGNVSTGNGISNVNLEEQFK	780							
Qy	781	VRNTDDIKACGMAIGD	840	QSWNPNKYKILGKANKNIGISKTANGSK	840							
Db	781	VRNTDDIKACGMAIGD	840	QSWNPNKYKILGKANKNIGISKTANGSK	840							
Qy	841	NGGNTNLPTNTTNSA	900	RNNALQNAFPQPSATPNLVAINQHDFG	900							
Db	841	NGGNTNLPTNTTNSA	900	RNNALQNAFPQPSATPNLVAINQHDFG	900							
Qy	901	DIDTLYNSGAQGRDL	960	LOTLLIDSHDAGYARKMIDATSAEITK	960							
Db	901	DIDTLYNSGAQGRDL	960	LOTLLIDSHDAGYARKMIDATSAEITK	960							
Qy	961	HKTSGLQTLISNAMIL	1020	NSRLVNLRSRHTNHDSPAKRLQAKQ	1020							
Db	961	HKTSGLQTLISNAMIL	1020	NSRLVNLRSRHTNHDSPAKRLQAKQ	1020							
Qy	1021	APKYKEPTNNWANAIG	1080	TGTSNLNGSNASLYGTSGAVDAYLNG	1080							
Db	1021	APKYKEPTNNWANAIG	1080	TGTSNLNGSNASLYGTSGAVDAYLNG	1080							

Db 1021 APKYEKPTNWNANATGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSSFPNN
 Qy 1081 RANSLNSGANNNTFCGYVSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYHYLA 1
 Db 1081 RANSLNSGANNNTFCGYVSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYHYLA 1140
 Qy 1141 YSAATRASGYGDFAFRFRNALVLKPSVGYVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
 Db 1141 YSAATRASGYGDFAFRFRNALVLKPSVGYVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
 Qy 1201 SANVEARYYGDTSYFYMNAGVLQFEAFHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
 Db 1201 SANVEARYYGDTSYFYMNAGVLQFEAFHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
 Qy 1261 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1296
 Db 1261 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1296
 RESULT 2
 US-08-471-491-3
 ; Sequence 3, Application US/08471491B
 ; Patent No. 6090611
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; APPLICANT: Bugnoli, Massimo
 ; APPLICANT: Telford, John
 ; APPLICANT: Macchia, Giovanni
 ; APPLICANT: Rappuoli, Rino
 ; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
 ; FILE REFERENCE: CHIR0044
 ; CURRENT APPLICATION NUMBER: US/08/471,491B
 ; CURRENT FILING DATE: 1995-06-06
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 1296
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 US-08-471-491-3
 Query Match 100.0%; Score 6671; DB 3; Length 1296;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MEIOOTHRKINRPLYSALVGALVSITPQSHAAFFTTVIIPAIIVGGTATGTAVGTVSGL 60
 Db 1 MEIOOTHRKINRPLYSALVGALVSITPQSHAAFFTTVIIPAIIVGGTATGTAVGTVSGL 60
 Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNKEYDLYRSLSSKIDGGWDMGNAR 120
 Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNKEYDLYRSLSSKIDGGWDMGNAR 120
 Qy 121 HWYKGGQONKLEVDMDKDAVGTYYTLISGLRNFETGGDLVDYVMOKATURLQGFNGNSTSYKD 180
 Db 121 HWYKGGQONKLEVDMDKDAVGTYYTLISGLRNFETGGDLVDYVMOKATURLQGFNGNSTSYKD 180
 Qy 181 SADRTTRVDENAKNISIDNFVEINNRVSGAGRKASSTVLTLOASEGTTSDKNABISLYD 240
 Db 181 SADRTTRVDENAKNISIDNFVEINNRVSGAGRKASSTVLTLOASEGTTSDKNABISLYD 240
 Qy 241 GATLNLASSSVKLMGNVWMGRLOQYVGAYLAPSYSTINTSKYTGVEVFNHFLTVDGKNAQA 300
 Db 241 GATLNLASSSVKLMGNVWMGRLOQYVGAYLAPSYSTINTSKYTGVEVFNHFLTVDGKNAQA 300
 Qy 301 GIIANKKNTIGTLDLWQAGLNIITAPPEGYKDKPNNTPSOGAKNDKNSAKNDKQESS 360
 Db 301 GIIANKKNTIGTLDLWQAGLNIITAPPEGYKDKPNNTPSOGAKNDKNSAKNDKQESS 360
 Qy 361 QNNSNTVOINPPNSAKTEVQPTQVIDGPFAGGKDTVYINRINTNADGTRVGGFKASL 420

Db 361 QNNSNTQVTPNPNSAQKTEVQTVIDGPFAGKDTVVNINRINTNADGTTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGVNLNSQAGSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGVNLNSQAGSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Qy 481 KAGTDTKNGTATFNDNDISLGRFVNLKVDATNFKGIDTNGGFGNTLDFSGVTDKVNINK 540
Db 481 KAGTDTKNGTATFNDNDISLGRFVNLKVDATNFKGIDTNGGFGNTLDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLPSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLPSGGV 600
Qy 601 KFKGGEKLVDFEYFSPWNYFDARINKVETNKLAFAFGQSPGWGTSKLMFNLLTGQNA 660
Db 601 KFKGGEKLVDFEYFSPWNYFDARINKVETNKLAFAFGQSPGWGTSKLMFNLLTGQNA 660
Qy 661 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 720
Db 661 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 720
Qy 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISVNVLEEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISVNVLEEQFKERLALYNNNRMDTCV 780
Qy 781 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 840
Db 781 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 840
Qy 841 NGGNTNLTNTTSNARSANNAALQAPPAQSPATPNLVA INOHDFGTIESVPELANRSK 900
Db 841 NGGNTNLTNTTSNARSANNAALQAPPAQSPATPNLVA INOHDFGTIESVPELANRSK 900
Qy 901 DTDLYANSAGORLLOQLLLDSDHAGYARKMIDATSAEITKQLTATTTLNNTASLE 960
Db 901 DTDLYANSAGORLLOQLLLDSDHAGYARKMIDATSAEITKQLTATTTLNNTASLE 960
Qy 961 HKTSGIQTLSLNSMLNRLNRRHTNHIDSFARLQALKDQKFALESAAEVLQF 1020
Db 961 HKTSGIQTLSLNSMLNRLNRRHTNHIDSFARLQALKDQKFALESAAEVLQF 1020
Qy 1021 APKYKPTNVWANAIGGTSNLNGSNASLYGTSGAVDAYLNGQVEAIVGGFGSYGSFNN 1080
Db 1021 APKYKPTNVWANAIGGTSNLNGSNASLYGTSGAVDAYLNGQVEAIVGGFGSYGSFNN 1080
Qy 1081 RANSLNGANNNTFNGYSRIFANQHEFDEAOGALGSDOSSLNFKSALLQDLNQSXYHLA 1140
Db 1081 RANSLNGANNNTFNGYSRIFANQHEFDEAOGALGSDOSSLNFKSALLQDLNQSXYHLA 1140
Qy 1141 YSAATRASGYDPAFRNALVLKPSVGSYVNLHLSNIGHFASNLGMRYSF 1200
Db 1141 YSAATRASGYDPAFRNALVLKPSVGSYVNLHLSNIGHFASNLGMRYSF 1200
Qy 1201 SANVEARYYGTSTFYMAGVQLQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
Db 1201 SANVEARYYGTSTFYMAGVQLQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
Qy 1261 LKLAKEVFLNGLGVVYVHLNLSNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNGLGVVYVHLNLSNIGHFASNLGMRYSF 1296

RESULT 3
US-08-466-662-3
; Sequence 3, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino

; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-3

Query Match 100.0%; Score 6671; DB 4; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MEIQOTHRKINRPLVSLALVGVLSITPQSHAAAFETTIIIPAIIVGGIATGAVGTVSGL 60
Db 1 MEIQOTHRKINRPLVSLALVGVLSITPQSHAAAFETTIIIPAIIVGGIATGAVGTVSGL 60
Qy 61 LSHGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGQWDMGNAAR 120
Db 61 LSHGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGQWDMGNAAR 120
Qy 121 HWYVGGQONKLEVDKMDAVGTYTTLGLRNFRTGGDLVDNMOKATRLRGOFNGNSFTSYKD 180
Db 121 HWYVGGQONKLEVDKMDAVGTYTTLGLRNFRTGGDLVDNMOKATRLRGOFNGNSFTSYKD 180
Qy 181 SADRTTRVDNFNAKNSIDNFEVNNRVSAGRKASSTVLTILQASBGITSDKNAEISLYD 240
Db 181 SADRTTRVDNFNAKNSIDNFEVNNRVSAGRKASSTVLTILQASBGITSDKNAEISLYD 240
Qy 241 GATNLASSSVKLMGVNMGRLQYVGYAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGVNMGRLQYVGYAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Qy 301 GIANKTKNIGTLDWOSAGLNIIPAPPEGYKDKPNPTPSQCAKNDKNSAKNDKQESS 360
Db 301 GIANKTKNIGTLDWOSAGLNIIPAPPEGYKDKPNPTPSQCAKNDKNSAKNDKQESS 360
Qy 361 QNNSNTQVTPNPNSAQKTEVQTVIDGPFAGKDTVVNINRINTNADGTTIRVGGFKASL 420
Db 361 QNNSNTQVTPNPNSAQKTEVQTVIDGPFAGKDTVVNINRINTNADGTTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGVNLNSQAGSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGVNLNSQAGSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Qy 481 KAGTDTKNGTATFNDNDISLGRFVNLKVDATNFKGIDTNGGFGNTLDFSGVTDKVNINK 540
Db 481 KAGTDTKNGTATFNDNDISLGRFVNLKVDATNFKGIDTNGGFGNTLDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLPSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLPSGGV 600
Qy 601 KFKGGEKLVDFEYFSPWNYFDARINKVETNKLAFAFGQSPGWGTSKLMFNLLTGQNA 660
Db 601 KFKGGEKLVDFEYFSPWNYFDARINKVETNKLAFAFGQSPGWGTSKLMFNLLTGQNA 660
Qy 661 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 720
Db 661 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 720
Qy 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISVNVLEEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISVNVLEEQFKERLALYNNNRMDTCV 780
Qy 781 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 840
Db 781 VMDYSQFSLTIOGDFINNOGTTINYLVRGKGVATISVNVLEEQFKERLALYNNNRMDTCV 840

QY 841 NGGNTNLPTNTTNSARSANNALQAAPPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Db 841 NGGNTNLPTNTTNSARSANNALQAAPPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
QY 901 DIDTLYANSQAQRDLLOTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Db 901 DIDTLYANSQAQRDLLOTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
QY 961 HKTSGLOTLSLSNAMILNSRLVNSRHTNHDIDSFARKLQALKDQKFALESAAEVLVYQF 1020
Db 961 HKTSGLOTLSLSNAMILNSRLVNSRHTNHDIDSFARKLQALKDQKFALESAAEVLVYQF 1020
QY 1021 APKYEKPNTWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
Db 1021 APKYEKPNTWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
QY 1081 RANSLNGANTNFVYSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYYHLA 1140
Db 1081 RANSLNGANTNFVYSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYYHLA 1140
QY 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
QY 1201 SANYEARYYYGDTSYFYMNAGVLQEFARHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1201 SANYEARYYYGDTSYFYMNAGVLQEFARHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
QY 1261 LKLAKEVFLNLGVVYLNHLISNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNLGVVYLNHLISNIGHFASNLGMRYSF 1296

RESULT 4

US-08-200-232-2
; Sequence 2, Application US/08200232
; Patent No. 5721349
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,232
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-200-232-2

Query Match 94.5%; Score 6303.5; DB 1; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;
QY 1 MEIQOTHRKNRPLVSLALVGLYSITPQOSHAAFFTTVIPAIVGGIATGTAAGTYSGL 60
Db 1 MEIQOTHRKNRPLVSLALVGLYSITPQOSHAAFFTTVIPAIVGGIATGTAAGTYSGL 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGWAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGWAAR 120
QY 121 HYWVGKQONKLEVDMDKAVGTYTLISGLRNFPTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HYWVGKQONKLEVDMDKAVGTYTLISGLRNFPTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
QY 181 SADRTTRVDENAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSSDKNAEISLYD 240
Db 181 SADRTTRVDENAKNILIDNFEINNRVSGAGRKASSTVLTLOASEGITSSDKNAEISLYD 240
QY 241 GATLNLASSSVKLMGNVWMLQVYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATLNLASSSVKLMGNVWMLQVYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
QY 301 GIANKKNTNIGTLDWOSAGLNIIPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
Db 301 GIANKKNTNIGTLDWOSAGLNIIPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
QY 361 QNNSNTQVIPPNSAOKTEVQPTQVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVIPPNSAOKTEVQPTQVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
QY 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANPEF 480
Db 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANPEF 480
QY 481 KAGDTKNGTATFNNDISLGRFVNLKVDATANKGIDTGNNGENTLOFSVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDATANKGIDTGNNGENTLOFSVTDKVNINK 540
QY 541 LITASTNVAVKFNINELIVKTNISVGEYTHFSEDIQSOSRINTVRLTGTSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKTNISVGEYTHFSEDIQSOSRINTVRLTGTSLFSGGV 600
QY 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNLTGONA 660
Db 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNLTGONA 660
QY 661 VMDYSQFSLTIQDGFINNQCTINYLVRGKQVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLTIQDGFINNQCTINYLVRGKQVATLSVGNAAAMFNNDIDSATGFYKPLI 720
QY 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV 780
QY 781 VRNTDDIKACGMAIGDQSWNNPNYKYLIGKANKNTGISKTANGSKISVYILGNSTPTE 840
Db 781 VRNTDDIKACGMAIGDQSWNNPNYKYLIGKANKNTGISKTANGSKISVYILGNSTPTE 840
QY 841 NGGNTNLPTNTTNSARSANNALQAAPPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Db 841 NGGNTNLPTNTTNSARSANNALQAAPPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
QY 901 DIDTLYANSQAQRDLLOTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Db 901 DIDTLYANSQAQRDLLOTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
QY 961 HKTSGLOTLSLSNAMILNSRLVNSRHTNHDIDSFARKLQALKDQKFALESAAEVLVYQF 1020
Db 961 HKTSGLOTLSLSNAMILNSRLVNSRHTNHDIDSFARKLQALKDQKFALESAAEVLVYQF 1020

Db 952 HKTSSLQTLSSLNAMIINSLVNLRRHTNNIDSFARLQALQDQFASLESAAEVLQF 1011
QY 1021 APYKPTNWNANAIGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1012 APYKPTNWNANAIGGASLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1071
QY 1081 RANSLNGANNTNFGVYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYYLA 1140
Db 1072 QANSLNGANNTNFGVYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYYLA 1131
QY 1141 YSAATRASVGYDFAFERNALVLKPSVGVSYNHLGTSNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1132 YSAATRASVGYDFAFERNALVLKPSVGVSYNHLGTSNFKSNSTNKVALSNGSSQHLFNA 1191
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGE 1260
Db 1192 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGE 1251
QY 1261 LKLAKEVFLNLGVVYVYHNLISNIGHGASNLGMRYSF 1296
Db 1252 LKLAKEVFLNLGVVYVYHNLISNIGHGASNLGMRYSF 1287

RESULT 5
PCT-US95-02219-2
; Sequence 2, Application PC/TUS9502219
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSES: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02219-2

Query Match 94.5%; Score 6303.5; DB 5; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;

QY 1 MEIQOQTHRKINRPLVSLALVGALVSTIPQOSHAAFTTIIIPAIIVGGIATGTAVGTVSG 60
Db 1 MEIQOQTHRKINRPLVSLALVGALVSTIPQOSHAAFTTIIIPAIIVGGIATGTAVGTVSG 60
QY 61 LSNGLKQAEPAKTPDKPVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120

Db 61 LGWGLKQAEPAKTPDKPVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
QY 121 HWYKGGQGNKLEVDKMDAVGTYTTLGSLNFTGGDLVDNMQKATLRLGFGNFSFTSYKD 180
Db 121 HWYKGGQGNKLEVDKMDAVGTYTTLGSLNFTGGDLVDNMQKATLRLGFGNFSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASRGITSDKNAEISLYD 240
Db 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASRGITSSKNAEISLYD 240
QY 241 GATLNLASSSVKLMGNVMGRLOYVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 GATLNLASSSVKLMGNVMGRLOYVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
QY 301 GIITANKKTITGLDQWOSAGLNIAPPEGGYKDKPNMTPSOCAKNDKNEKAKNDKQESS 360
Db 301 GIITANKKTITGLDQWOSAGLNIAPPEGGYKDKPNMTPSOQSG-----AKNDKQESS 352
QY 361 QNNSNTQVINPPNSAOKTEVQPTQVIDGPFAGGKDFVNNIRINTNADGTIRVGGFKASL 420
Db 353 QNNSNTQVINPPNSTOKTEVQPTQVIDGPFAGGKDFVNNIRINTKADGTIKVGGFKASL 412
QY 421 TTNAAHHTGKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Db 413 TTNAAHHTGKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 472
QY 481 KAGDTKNGTATFNNDISLGRFVNKLVDAHTANFKGIDTGNGGFWNTLDFSGVTDKVNINK 540
Db 473 KAGVDTKNGTATFNNDISLGRFVNKLVDAHTANFKGIDTGNGGFWNTLDFSGVTNKVNINK 532
QY 541 LITASTNVAVKNFNINELIVKTNIGSVGEYTHFSEIDIGSQSRINTVRLTGTFRSLPSGGV 600
Db 533 LITASTNVAVKNFNINELIVKTNIGSVGEYTHFSEIDIGSQSRINTVRLTGTFRSLPSGGV 592
QY 601 KFKGGEKLVDFEYFSPWNYFDARNIKNVEITNKLAFPGQSPGWGTSKLMFNLLTGQNA 660
Db 593 KFKGGEKLVDFEYFSPWNYFDARNIKNVEITNKLAFPGQSPGWGTSKLMFNLLTGQNA 652
QY 661 VMDYSQFSLTTOGDFINNOGTINYLVRGKVTATLSVGNAAAMFNNDTSATGFYKPLI 720
Db 653 VMDYSQFSLTTOGDFINNOGTINYLVRGKVTATLSVGNAAAMFNNDTSATGFYKPLI 712
QY 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 713 KINSADLLIKNTEHVLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 772
QY 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIGRAWKNGIGSKTANGSKISVYVLGNSTPTE 840
Db 773 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIGRAWKNGIGSKTANGSKISVYVLGNSTPTE 832
QY 841 NGGNTNLTPTNTSNARSANALAQNAQPAQSAFNPALVAINQHDGTTIESFELANRSK 900
Db 833 NGGNTNLTPTNTNNARFASYALIKNAPFAH-SATPNLVAIINQHDGTTIESFELANRSK 891
QY 901 DIDTLYANSAGQGRDLLQTLIDSDHAGYARKMIDATSAEITKOLNTATTLNNTIASLE 960
Db 892 DIDTLYANSAGQGRDLLQTLIDSDHAGYARTMIDATSAEITKOLNTATTLNNTIASLE 951
QY 961 HKTSSLQTLSSLNAMIINSLVNLRRHTNNIDSFARLQALQDQFASLESAAEVLQF 1020
Db 952 HKTSSLQTLSSLNAMIINSLVNLRRHTNNIDSFARLQALQDQFASLESAAEVLQF 1011
QY 1021 APYKPTNWNANAIGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1012 APYKPTNWNANAIGGASLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1071
QY 1081 RANSLNGANNTNFGVYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYYLA 1140
Db 1072 QANSLNGANNTNFGVYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYYLA 1131
QY 1141 YSAATRASVGYDFAFERNALVLKPSVGVSYNHLGTSNFKSNSTNOVALKNGSSQHLFNA 1200

```

Db 1132 YSAATRASYGYDFAFPRNALVLKPSVGVSYNHLGSTNFKSNSTNKKVALSNGSSSQHLFNA 119
QY 1201 SANVEARYYYGDTSYFYMNAGVLQBFARHVGSSNNAASLNTFKVNAARPLNTHARVMMGGE 1260
Db 1192 SANVEARYYYGDTSYFYMNAGVLQBFARHVGSSNNAASLNTFKVNAARPLNTHARVMMGGE 1251
QY 1261 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1296
Db 1252 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1287

RESULT 6
PCT-US95-02219A-2
; Sequence 2, Application PC/TUS9502219A
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Tummuru, Murali KR
; APPLICANT: Cao, Ping
; APPLICANT: Thompson, Stuart A.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02219A-2

Query Match 94.5%; Score 6303.5; DB 5; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps

QY 1 MEIQOTHRKNRPLVSLALVGALVSITPQSHAAFFTTVIIPIAVGGIATGATGVTVSG 60
Db 1 MEIQOTHRKNRPLVSLALVGALVSITPQSHAAFFTTVIIPIAVGGIATGATGVTVSG 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDWGNAR 120
Db 61 LGWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDWGNAR 120
QY 121 HYWKKGQONKLEVDKMDAVGYTTLSGLRNFETGGDLDMOKATLRLCQFNGNSFTSYKD 180
Db 121 HYWKKGQONKLEVDKMDAVGYTKLSGLRNFETGGDLDMOKATLRLCQFNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFEVNNRVGSGAKRKAASVTVLQASEGJTSKNAEISLDY 240

```

181	Db	SADRTTRVDFNAKNILIDNPLNINRVSGAGRKASSTVFLVLAQSEGTSSKNAEISLYD	240
241	Qy	GATLNLASSSVKLMGNVWGRQYQYVAYLAPYSYSTINTSKVTGEVFNHLTVGDKNAQA	300
		: : : : : : : : :	
241	Db	GATLNLASSSVKLMGNVWGRQYQYVAYLAPYSYSTINTSKVTGEVFNHLTVGDHNAQA	300
301	Qy	GIITANKKTIGTLDLWQSAGLNIIPAPPEGYKDKPNNTPSQSGAKNDKNAESAKDKQESS	360
		: : : : : : : : :	
301	Db	GIITASNTKTHIGTLDLWQSAGLNIIPAPPEGYKDKPNNTPSQSG-----AKNDKQESS	352
361	Qy	QNNSTQVINPPNSAOKTEVQPTQYIDGPPFAGGKDTVYNNIRNTNAOGTIRVGGFKASL	420
353	Db	QNNSTQVINPPNSTOKTEVQPTQYIDGPPFAGGKDTVYNNIRNTKADGTIKVGGFKASL	412
421	Qy	TTNAAHLHGKGVNLSNOASGRSLVENLTCNITVDGPLRVNNQGVGYALAGSANPEF	480
		: : : : : : : : :	
413	Db	TTNAAHLNLTGKGVNLSNOASGRTLLENLTCNITVDGPLRVNNQGVGYALAGSANPEF	472
481	Qy	KAGCTDKNGCTATFNNDISLGRFVNLKVDAAHTANFKGIDTGGNGFNTLDFSGVTDKVNINK	540
		: : : : : : : : :	
473	Db	KAGVDTKNGCTATFNNDISLGRFVNLKVDAAHTANFKGIDTGGNGFNTLDFSGVTNKVNINK	532
541	Qy	LIITASTNVAVKNFINELIVKTINGISVGEYTHFSEDIQSQRINTVRLTGTGRSLFSGGV	600
533	Db	LIITASTNVAVKNFINELIVKTINGVSGEYTHFSEDIQSQRINTVRLTGTGRSIFSGGV	592
601	Qy	KFKGGKGLVJDEYFYSPPWNYFDARNIKNVEITNKLAFGPQSGPMGCTSKLMFNNTLQONA	660
		: : : : : : : : :	
593	Db	KFKSGKGLVJDEYFYSPPWNYFDARNIKNVEITRKFPASSTPENPGCTSKLMFNNTLQONA	652
661	Qy	VMDYSQFSNLTIQGDFINNOGTINYLVRGGKVATILSVGNAAAMFNNDIDSATGYFKPLI	720
		: : : : : : : : :	
653	Db	VMDYSQFSNLTIQGDFINNOGTINYLVRGGKVATILVGNAAAMFNNDIDSATGYFKPLI	712
721	Qy	KINSAQDLKNTHEVHLKAKIIGYGNVSTGTNGISNVALBEOQFERLALYNNNRMDTCV	780
713	Db	KINSAQDLKNTHEVHLKAKIIGYGNVSTGTNGISNVALBEOQFERLALYNNNRMDTCV	772
781	Qy	VRNTDDIKACGMAIGQSMVNPNDNYKYLLIGKAWKNIGISTANGSKISVYVYLGNSTPTE	840
		: : : : : : : : :	
773	Db	VRNTDDIKACGMAIGQSMVNPNDNYKYLLIGKAWKNIGISTANGSKISVYVYLGNSTPTE	832
841	Qy	NGCNTNLTPTNTTNSARSANNAALQAPASATPNLVAINQHDFGTIESVFELANRSK	900
		: : : : : : : : :	
833	Db	NGCNTNLTPTNTTNNAREASALIKNAPEAH-SATPNLVAINQHDFGTIESVFELANRSK	891
901	Qy	DIDTLVANSQAQRDLLOTLILLDSDHAGYARKWIDATSANETIKOLNTATTTLNNIASLE	960
892	Db	DIDTLVANSQAQRDLLOTLILLDSDHAGYARTMIDATSANETIKOLNTATTTLNNIASLE	951
961	Qy	HKTSGLOTLISLNSAMTLNSRLNLSRRHHNHDSEAKRLQALKDQKFASLESAAEVLQYF	1020
		: : : : : : : : :	
952	Db	HKTSSLOTLISLNSAMTLNSRLNLSRRHTNNIDSFAKRLQALKDQKFASLESAAEVLQYF	1011
1021	Qy	APKYEXPTNVWANAIGGTSLLNNGSNASLYGTSGAYDAYLNGQVEAIVGGFGSGYGSFPNN	1080
		: : : : : : : : :	
1012	Db	APKYEXPTNVWANAIGGASLLNNGGNASLYGTSGAYDAYLNGQVEAIVGGFGSGYGSFPNN	1071
1081	Qy	RANSLNSGANNTHGVYSRIFANQHEFDEPAQOAGLGSQOSSLNFKSALLQDLNOSYHYLA	1140
		: : : : : : : : :	
1072	Db	QANSLNSGANNTHGVYSRIFANQHEFDEPAQOAGLGSQOSSLNFKSALLRDLNOSYHYLA	1131
1141	Qy	YSAATRASGYDFAFFRNALVLKPSVGVSYNHLGTFNFKSNSTNOVALKNGSSSOHLFNA	1200
		: : : : : : : : :	
1132	Db	YSAATRASGYDFAFFRNALVLKPSVGVSYNHLGTFNFKSNSTNKVALNSGSSSOHLFNA	1191
1201	Qy	SANVEARYYYGDTSYEYFMNAGVLQEFARHVGNNAAASLNTFKYNAARNPLNTHARYMMGGE	1260
		: : : : : : : : :	
1192	Db	SANVEARYYYGDTSYEYFMNAGVLQEFANFGSSNAVSLNTEFKVNAARNPLNTHARYMMGGE	1251
1261	Qy	LKLAKEVFLNLGVVYVYUHNILNISIGHFASNLGMRYSE	1296
		: : : : : : : : :	
1252	Db	LKLAKEVFLNLGVVYVYUHNILNISIGHFASNLGMRYSE	1287

[illegible]

CLASSIFICATION: 435
PRIOR APPLICATION DATA: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2

Query Match 4.8%; Score 320.5; DB 1; Length 1612;
Best Local Similarity 18.7%; Pred. No. 3.8e-13;
Matches 293; Conservative 186; Mismatches 555; Indels 533; Gaps 69;

QY 20 VGLVSIITPQSHAFPTTIIIPAIVGGIATGTAFTVSGLLSGWGLQAEAEANKTPDKPD 79
DB 77 LGFTITIAQO---AKFFNFV-----AAGKILNITGGI-TVQEAASNTINAQN 120
QY 80 KVRRIQAGKGFN-----EPP-----NKEYDLYRSLSSKIDGGDW 115
DB 121 ALTKVHGGAANDLSGLSITFAVCPVLEFNLINPINSRSSYHLVSNRSKI----- 174
QY 116 GNAARHWYKGGQONKLEVDMDKDAVGYTLISGLRNFTHGGDLVNMOK-----ATLRGLQFN 171
DB 175 -----VNGG-----NGILNITNGFIQVSDNTFAGIKTINIDDCQ 208
QY 172 GNSFTYKDSADRTVRDFAKNKISI-DNFVEINNRVGG-----AGRKASSTVLTQAS- 225
DB 209 GLMFSTPDAA-----NTLQAGGNTINFGIDGTGKLVSKNGAATFNFVTGTL 260
QY 226 ---EGITSKDNAEISLYDGATLNLASSVKLMG-----NVWGRLOQVVGAY 268
DB 261 GGNLKGIIELNTHAIVA---GKLISLGAANAVIGTONGAGRAAGFIVSDNGNAATISGQ 317
QY 269 LAPSYTINTSKYTGVEVFNHL-----TVGDKNAAQAAGIANK-----KTNIGTLDLWQ 317
DB 318 VYAKNVVIOQANAGQVTFEHIIVDVLGGTTNEFTADSKVLIITENSFNGSTNFGNLD--- 374
QY 318 SAGLNIAPP-----EGGYKDKPNTPSQG-----AKNDKNESAKNDKQESSQ 361
DB 375 ---TQIVVPDTKILGNFIDGVKNNGNTAGVITFNANGALVSASTDPNIAVTNINAIAE 431
QY 362 NNSNTQV-----INPPNSAQKTEVQPTQVDPGPAAGKDTVVNINRINTN---ADGTI 411
DB 432 GAGVVELSGIHIELRLGNGGSTFKLADGTVINGP-----VNQALMNNNALAAGSI 483
QY 412 RVGGFKASLTNNAHLHIGKGVN-----LSNQAS-----GRSLIVENLTGNITVD- 457
DB 484 QLDG-SAITGD-----ICNGGVNAAALQHTITLANDASKILALDGANIIGANVGGAITHFOA 537
QY 458 --GPLRVNQVGYALAGSANEFFKAGTD-----TKNGTATFN----- 494
DB 538 NGGTIKLTNTQNNIV-----NFDLDTTDTKTVGDVASSLTNTQTLTNGISGTVVANTK 592
QY 495 -----NDISLGRV---NLKVDATHANF---KGIDTNGGFWNTLDFSGV 532
DB 593 TLAQLNIGSKTILNAGDVAINELVNIENSGVOLNHNHTYLIITKTINAAGOG-----QIIVA 648
QY 533 TDKVINKLITASTNV-----AVKFNENINELI-----VKTINGISVG 568
DB 649 ADPLNTNTTLADGTNLGSAENPLSTHFKATKAANDSILNVGKVNLYANNITNDANVG 708
QY 569 EYTHF-----SEDIGSQS---RINTVRLETGTSLFSGGVKFGKGEKLVIDEFYSPW 618

DB 709 S-LHFRSGGTSIVSGTVGGQGHKLNNLILDNCTTVKFLGDDTTFNGGKTIEGSKILQISN 767
QY 619 NYF-----DARNIKNVEITN---KLAFGPGSGPWGTSK-----LMFN---NLTGL 657
DB 768 NYTTDHVESADNTGTLEFVNTDPTITLNKQGAIFYGLVKQVLIISGPGNIVFNEIGNGV 827
QY 658 QNAVMDYSQFSLTIQDFINNOGTINYLVRGKGKATLSVGNAAAMFNNDIDISATGFYK 717
DB 828 HGTAANSISFENASL-----GTSIFLPSGTPDLVLTKISVG---NGVDN---FNA 873
QY 718 PLIKINSAODLIKNTHEHVLKAKIIG-----YGNVSTGTNGISNVNLEE 761
DB 874 PIVVSGIDSMINNGQIIGDKNIIALSGLSDNSITVNANTLYSGIRTTKNNQGTVTLSG 933
QY 762 QPKER-----LALYNNNRMD---TCVVRNTDDIKACGMAIGDQ----- 797
DB 934 GMPNPCTIYGLGLENGSPKLGKQVTFDTDYNNLGSIIANNVTINDVTLTGTGIAGTDFD 993
QY 798 -----SWVANNPDNYKYLIGKAWKNIGISKTAGSKISVYVLGNS-----TPTEN-- 841
DB 994 AKITLGSVNGNANVRFDSTFSPRSMIVATQANKGTIVYLGNALYSNIGSLDTPVASVR 1053
QY 842 -----NLPPTNTTSNARSANALQAAPPAPSAT-----GGNTT----- 846
DB 1054 FTGNDSCAGLOGNIYSONIDFGTYNLTILNSNVLGGTTAINGEIDLTLNLIIFANGTS 1113
QY 847 ---NLPPTNTTSNARSANALQAAPPAPSAT----- 875
DB 1114 TWGDNSTISITLNVSSNGIGQVVIADQAVNATTTGTTIKIODNANANFSGTQAVTLIQ 1173
QY 876 -----PNLVAINOHDFTGIESVFE-----LANRSKD-----IDTLYANSQ-AQGR 914
DB 1174 GGARENGTLGAPNFAVTGNSIFVKYELIRDSNODYVLTRTNDVLNVTTAVGNSAIAANAP 1233
QY 915 DLLQTL---LIDSHDAGYARKMID-----ATSANEITKOLNTATTTLNINIALEHKTSG 965
DB 1234 GVSQNSRCLSESTNTAAYNNMLLAKDPDVATEFVGAIATDTSAAVTVN---LNDTKOTQD 1291
QY 966 LQTLSLSANMILNSRLNLSRRHTNHDISFAKRLQALQDKQFASLESAAEVLVQFAPKYE 1025
DB 1292 L-----LSNRL---GTYRLSNAETSDVAGSATGAVSGDEAEVSYGVWAKPFFNIAEQDK 1344
QY 1026 KPTNVWANAIGTSLNNGSNASLYGTSGAVD---AYLNGQVEAIVG----- 1068
DB 1345 K-----GGIA---GYKAKTTGVVVGDLTTLASDNLMIGAAGITKTDIKHQDYKKGD 1392
QY 1069 --CGFGYGYSSFNRRANSNLNGANNTNFVYSRIFANQHEFDEAOGALGSDOSSLNFKS 1126
DB 1393 KTDINGLSFLSGYSGQOLVKNFNFAQGNISFTLNKVKSKSQRYFFESNGKMSKQIAAGNYD- 1451
QY 1127 ALLQDLNOSVHYLAYSAAATRASVGYDFAPFRNALVLKPSVSVNHLGTSNFKSNSTNOV 1186
DB 1452 -----NMVF-----GGNLIIFYDYNAMPNVLV-TPMAGLSYLLKSSNENYKETGTT-V 1496
QY 1187 ALKNGSS 1193
DB 1497 ANKRINS 1503

RESULT 11
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; Bldg. 1

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 4.4%; Score 292; DB 4; Length 1529;

Best Local Similarity 20.0%; Pred. No. 3e-11;
Matches 261; Conservative 161; Mismatches 482; Indels 402; Gaps 54;

QY 134 VDMK-----DAVGTYTSLGLRNFVG-----GO-----LDVNNQKATL 165
DB 360 VDAKEWLLDPDDVSIETLTSGRNNTGENOGYTTGDGTEKSPKGNISKPTLTNSTLEQL 419
QY 166 RLQGF-----NGNSFTSY-----KQADRTTRVDENAKNISIDNEVE 202
DB 420 RRGSVNITANNRIYVNSINLSGLTLTKRDGVKINGDITSNENGL-TIKAGSWD 478
QY 203 INRVGSGAG-----RKASSTVLTLOASEGITSOKNAEISLYDGAT 243
QY 479 VHKNTLTGTGFLNIVAGDSVAFEREGDKARNATDAQITAGGT--ITVKKDDKQFRFNVS 536
QY 244 LNLASSSVKLMGN-----VWGRLOYVG-----AYLAPSYSTIN 277
DB 537 INGTGKGLKFIANNFTHKFGELNISIVTINOTTKDKVYWNASKDSYWNVSLSLTIN 596
QY 278 TSKVTGEVNF-----NHLTVGDKNAQAAGIIAN-----KKTNIGTLDLWQAGLNIITAPPEG 329
DB 597 TVQKFTFIKFDVSGNSQDLRSRRSFAGVHENGIGKTN-----FNIGANAKA 645
QY 330 GYKDKPNTTPOSGAKNDKNSAKNDKQF-----SSQNSNTQVNPNSAQKTEV 380
DB 646 LFLKLPN-----AATDPKELPTITFNANITATGNSDSSVMFDIHANLTSRA 691
QY 381 OPTQVIDGPEAGKDTVVNINRINTNA-----DGTIRVGGFKASL--TTNAHLHIGKG 433
DB 692 AGINWDSNITGGLDFSTSHNRNNAEIEKKDLTINATGNSFSLKQTKDSFYNEYSRHA 751
QY 434 VNLSNQASGRSLIVENLT---GNITVDGPLRVNNVGGYALAGSSANFEFFAKGTDTKNGT 490
DB 752 INSSH-----NLTLGGNVLGGE--NSSSSITGNINITNKANVTLQADTSNSN-T 799
QY 491 ATFNNDISLGRV-----NLKVDHAHTANFKIDGTGNGGFFNLDPSGVTDKVNINKLITAS-T 546

DB 800 GLKKRRLTLGNISVEGNLSLTGANANIV-----GNLSIAEDSTFKGEASDNLNITGTFT 853
QY 547 NVAVKFNELIVKNTGINSIGVEYTHFSEDSQSQRINTVRLTGT-RLSLFSGVGFKGG 605
DB 854 NNGTANINQGVKLOG-----DINKGGLNITTNASGTOKTIIINGNITNEKG 902
QY 606 EKLVIDEYFYPWNYFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNTLGNVMDYS 665
DB 903 D-----LNINKIKADAELIQG-----GNIS-----Q 923
QY 666 QFSNLTIOGDFIN--NOGTINYLVRGKVATLSVGNAAAMFNNDDISA-----TGFIKP 718
DB 924 KEGNLTISSDKVNITNQITIKAGVEGRSDSSEAENALTIQTKELKLAGDLNISGFKA 983
QY 719 LKINSAQDL-INKTEHVLLKAKIIGYGV---STGTNGISVNVNLEQPKERLALYNNN 774
DB 984 EITAKNGSDLTIGNAGSGGNADAKKVTDFKVKDSKISTDG--HNVTLSSEKVTSGNSNAGN 1042
QY 775 RMDTCVVRNTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKNIGISKTAGNSKISVYILG 834
DB 1043 DNSTGLTISAKDV-----TVNN-----NVTSHKTINISAAA---G 1074
QY 835 NSTPTENGNGTNNLPTN---TTSNARSANNAALAAQAPSAFPATPNLVAINOHDFGTIE 890
DB 1075 NVTTKE--CTTINATTGSEVETAQNGTIRGNITSQNVTV---TATENLVTTENAVINATS 1129
QY 891 SVFELANRSKDI-----DLY 906
DB 1130 GTVNIKTGDIKGGIESTSGNVNITASNTLKVSNITGODVTVTADAGALTTTGTSTIS 1189
QY 907 ANSG-----AQGRDLI-----OTLLTDSHDAGYARKMID 935
DB 1190 ATTGNANITTKTDINGKVESSGSVTLVATGATLAVGNISGNTVITA-DSKLTSTVG 1248
QY 936 AT-----SANEITKOLNTATTTLLNNIASLEHKTSGLOTLSLSNAM 975
DB 1249 STINGTNSVTSSQSDIEGTISGNTVNVYTAGTDLTIGNSAKVEAK--NGAATLTAESGK 1307
QY 976 ILNSRLVNLRRHTNHIDSFARKLQAKDKFASLESAAEVLQFAPKYEKPTNVANAI 1035
DB 1308 L-----TTQCGSSITSSNGOTTAKDSSIAGTNAANVTLTNTGTTTGTGDSKINAT 1360
QY 1036 GGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFSGSYGSFSP-----NRANSLNSGAN 1090
DB 1361 SGTLTINAKDALDGAASGDRTVVNNATNAGSGNVTAKTSSSVNITGDLNTINGLNIIE 1420
QY 1091 NTNFGVYSRIFANQHEFDFE-AQAGLGDSSSLNFKSAL--LQDLNOSYHYLAISAATRA 1147
DB 1421 NGRNTVRLR---GKEIDVKYIQPGVASVEEVEIAKRVLEKVKDLSDE-----ERETLA 1470
QY 1148 SYGYDFAFFRNALVLKPSGVSYNHLGSTNFKSNSTNOVALKNGSS 1193
DB 1471 KLGVSAVRE-----VEPNNAITVN--TQNEFTTKPSSQVITISEGKA 1509

RESULT 13
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-728-470-9

Query Match 4.3%; Score 287.5; DB 2; Length 1338;
Best Local Similarity 20.3%; Pred. No. 4.9e-11;
Matches 272; Conservative 165; Mismatches 539; Indels 365; Gaps 58;

QY 10 INRPLVSLALVGLVITPQSHAAFFTTVIPAIVGGIATGATGTVSGLLSGLKQAE 69
DB 186 IEAPSASRVELGA-----DRNSHSAEVIKVTLLKNNSTLTLNT-TISNL----LKAH 235
QY 70 EAKTPDKPKVRIQAGKGFNEPNKEYDLYRLSLSSKIDGGWDGNAARHYYWKGQQ 129
DB 236 VVNITARRKLT-----NSSISIERG-----SHLILHSEGGQGG 270
130 NKLEVDKMDKAVGTVL-----SGLRNFTEGGDLVNM-----K 162
271 VOIDKQITSEGNLTIYSGWVDVHKHNTLGSGLNLTITTEGDIAPEDKSGRNNLTITAQ 330
QY 163 ATRLGQFNGNSEFT-----SYKDS-ADRTTRVDFNAK-----ISIDNFEVNN 205
DB 331 GTITSGNSNGFRFNNVSLNSLGGKLSFTSDREDRGRRTKGNISKNKFDGTLNIGTVDISM 390
QY 206 RVG--SCAGRKASST---VLTQASEGI-----TSKNAEIS--LYDQAT 243
DB 391 KAPKVSFYRDKGRTYWNVTTLNVTSGKFNLSIDTSGSGSTGSPSIRNAELNGITFNKAT 450
QY 244 LNLASSSVKLMGNVWMLQVVGAYLAPSYSTINTSKVT-----GEVNFNHLTVGDK 295
DB 451 FNIAQGTANFS-----IKASIPFFKSNANYALFNEDISVSGGSGVNFK-LNASS 500
QY 296 NAAQAGIANKTKTIGTLDLQWAGLNIAPPEGGYKDKPNTPSQSAKNDKRESAKND 355
DB 501 NIOTPGVII-KSQNFV-----SGSTLNLKAEQ-----STAFASIENDLNLATGG 547
QY 356 KQESSQNNSTQVNPNSQAQTEVQTPQIDGPFAGCKOTV-----VNINRI 403
DB 548 NITTRQEGTDSRVNKGVAANKNT-----FKGNTIFFGSAKATTEIKGNVTINK- 597
QY 404 NTNADGTRVGGFKASLTNTAAHLHICKGGVNLNQASGRSLIVENTGNITVDGPLRVN 463
DB 598 NTNA--TLR-----CANFAENKSPUNIACNVINNCNLTTAGSII--NIAGNLTVSKGANLQ 649

QY 464 NOVGGYALAGSSANFEK-AGTDTKNGTATFNNDISLGRFVNKVKVDAHTANFKGIDTNG 522
DB 650 -----AITNYTFNVAGSFDNNGAS-----NISIARGGAKFKDINN--- 684
QY 523 GFNTLDFSGVTDKVINIKLITASTNAVAKNFNINELIVKNGISVGEYTHFSEIDGQSQR 582
DB 685 -----TSSLNITNTSDTYRTIIR-----GNISKNKSGDLNIDK 718
QY 583 INTVRLGTGRSLFSGGVKFKGKELVIDEFYSPWNYFDARNIKNVEITNKLAF--GPO 640
DB 719 KSDAEIOIG-----GNISOKEGNLTISS-----KVNITNQITIKAGVE 757
QY 641 GSPWGTSKLMFNLTLCQNAVMDYSQFSNLTIOGDFINNOQTINYLVRGKVKVATLSYGNA 700
DB 758 GGRSDSSEAEANLTI-----QTKELKLAGD-LNISGFNKAEITAKNGSDLTIGNA 807
QY 701 AA-----MMFNNDI-----DSATGFYKPLIKNSA 725
DB 808 SGNADAKKVTDFKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNSTGL-----TISA 861
QY 726 QDLIKN--TEHVLLKAKIIGYGNVST-----GTNGISNVNLEEQFKERLALYNNNR 775
DB 862 KDVTVNNVTSKHTINISAAA-GNVTTKEGTTINATTSVEVTAQN-----GTIKGNIT 914
QY 776 MDTCVVRNTDDIKACGMAI-----GDQSMVNNPNKYKLIKAKWKNIGIKSTANGSKISV 830
DB 915 SQNVTVTATENLVTTENAVINATSGTVNISTKTGDKGIGIESTSGNVNITASGNTLKVSN 974
QY 831 YLGNSTPTENGNTNLP-----TNNTSNARSANNAQAOPAPQPSATPNLVAINQHDF 886
DB 975 ITGQDVTVTADAGALTTTAGSTISATTCGANITTKTGDKINGKVESSGSVTLVATG---- 1030
QY 887 GTIESVELANRSKIDITLVANSAGOCRDLLQTL-----LIDSHDAGYARKMIDAT-SAN 940
DB 1031 ----ATLAVGNISGNTVITADSGKLTSTVGSTINGNSVNTSSQSG-----DISGTISGN 1082
QY 941 EITKQLTATTTLNNIASLEHKTSGLOTLNLSNAMIINLSRLNLSRRHTNIDHIDFAKRLQ 1000
DB 1083 TVNVYTAGDGLTIGNSAKVEAK-NGAALTATAESCKL-----TTQTGSSITSSNGQTL 1134
QY 1001 ALKDOKFASLESAEVLVYQFAPKEYKPTNVWANAIGTSLNNGSNASLYGTSAGVDAYLN 1060
DB 1135 TAKDSSTAGNINAANVTLTGTLTTTGDSKINATSGTLTINAKDAKLDAAGSDRTVVN 1194
QY 1061 GOVEAIVGGFGSYGYSFN-----NRANSLNSGANNTNFGVYRIFANQHEFDE-AQA 1114
DB 1195 ATNAGSGNVYAKTSSSVNITGDLNTINGLNIISENGRNTVRLR-----GREIDVYIOPG 1250
QY 1115 LGSQSSSLNFKSAL--LQDLNQSYHYLAYSAATRASVGYDFAFPRNALVLKPSVGVSYNH 1172
DB 1251 VASVEEVIEAKRVLEKVKDLSD-----BRETAKLGVSAVRE-----VEPNNAITVN- 1298
QY 1173 LGSTNFKSNSTNQVALKNGSS 1193
DB 1299 -TQNEFTTKPSSQVITISEGA 1318

RESULT 14

US-08-719-641-9
Sequence 9, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Batenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia

		best boot similarity	20.96; Fred: NO: 4.9e-11;		
		Matches	272; Conservative	145; Mismatches	539; Indels
Qy	10	INRPLVSLALVGALVSITPQOSHAAFFTVIIIPAIIVGGIATGAVGTSGLLSWGLKQAE	69		Gaps
		: :			
Db	186	IEAPSASRVELGA----DRNSHSAEVIKYTKNNNTSLTLINT-TISNL---LKS AH	235		
		: :			
Qy	70	EANKTPDKPKWRIQAGKFNEFPNKEYDLYRSLSGSKIDGGWDGNGNAARHHVYWGQQ	129		
		: :			
Db	236	VVNITARRKLTV-----NSSISIERG-----SHLIHSEGGOGG	270		
		: :			
Qy	130	NKLEVDMDKDAVGYTYL-----SGLRNFTGGDLVNMQ-----K	162		
		: :			
Qy	271	VQIDKDITSEGNNITIYSGGVVDVHKHNITLGSGFLNITTKEGDIAPEDKSGRNNLTITAQ	330		
		: :			
Qy	163	ATLRGQFNGNST-----SYKDS-ADRTTRVDENAKN-----ISDNFEVEINN	205		
		: :			
Qy	331	GTITSGHSNGFRFNNSVLSIGLKSLTSDREDRGRTKGNISKNKFDGTLNISGTVDISM	390		
		: :			
Qy	206	RVG--SGAGRKA SST---VLTLQA SEGI-----TS DKNABIS--LYDGAT	243		
		: :			
Db	391	KAPKVSWFYRKGRTYWNVTTLNVNTSGSKENLSDTSGSGSTGPSIRNAELINGITENKAT	450		
		: :			
Qy	244	LNIASSVKLMGNVMGRLOYVGAYLAPSVYSTINTSKVT-----GEVNFNHLTVGDK	295		
		: :			
Db	451	FNIAQGSTANFS-----IKASIMPFKSNANYALFNEDI SVSGGGSYNFK-LNASSS	500		
		: :			
Qy	296	NAAQAGIIANKKTNI GTLDLWQSAGLN IAPPEGYKDKPNNTPSOSGAANDKEA SKND	355		
		: :			
Db	501	NIQT PGVII-KSQNFV-----SGGSTLNLKAE G-----STETAFSIENDLNLA TGG	547		
		: :			
Qy	356	KQESSQNNSNTQVNI PPNSAOKTEVOPTQVIDGPFAGGKDTV-----VINRI	403		
		: :			
Db	548	NITRQVEGTD SRV NKGVA AKKIT-----FKG GNTTFGSQKATTEIKGNVTINK-	597		
		: :			
Qy	404	NTNADGTRI VGGFKASLTITNA A HLHG KGVNLSNOASGRSLIVENITGNITVDGPL RVN	463		
		: :			

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 29.02 Seconds
(without alignments)
4291.240 Million cell updates/sec

Title: US-09-921:157-3
Perfect score: 6671
Sequence: 1 MEIQQTHRKINRPLVSLALV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6303.5	94.5	1287	2 B53739	vacuolating cytoto
2	6222.5	93.3	1291	2 S44983	vacuolating cytoto
3	6220	93.2	1290	2 G84630	vacuolating cytoto
4	6108	91.6	1288	2 E71884	vacuolating cytoto
5	1695.5	25.4	513	2 A53739	hypothetical vacuo
6	674	10.1	2902	2 C71953	toxin-like outer m
7	648.5	9.7	2893	2 A64556	toxin-like outer m
8	642.5	9.6	2529	2 B84635	toxin-like outer m
9	623	9.3	2399	2 H71879	toxin-like outer m
10	583	8.7	1943	2 B84596	toxin-like outer m
11	580.5	8.7	3194	2 D71917	toxin-like outer m
12	335	5.0	1655	2 E97835	hypothetical prote
13	314.5	4.7	1300	2 S07575	outer membrane pro
14	313	4.7	2249	2 A41477	190K surface anti
15	312.5	4.7	2020	2 C48399	ABC-type transport
16	310.5	4.7	2021	2 A97859	190-KDa cell surfa
17	310	4.6	1286	2 S28634	adhesin AIDA-I pre
18	310	4.6	1651	2 J21340	outer membrane pro
19	306	4.6	1643	2 D71630	outer membrane pro
20	303.5	4.5	2340	2 B71704	cell surface anti
21	297.5	4.5	1270	2 E85649	hypothetical prote
22	297.5	4.5	3705	2 A00123	probable autotrans
23	296.5	4.4	1268	2 B99789	hemagglutinin/hemo
24	291	4.4	1268	2 AB0204	conserved hypotet
25	291	4.4	1335	2 T17508	glycoprotein Vp260
26	291	4.4	1645	2 JN0896	crystalline surfac
27	290	4.3	1519	2 S41525	major ring-forming
28	283	4.2	1541	2 A37023	IgA-specific metal
29	282	4.2	1327	2 B90674	AIDA-I adhesin-lik

30	282	4.2	1349	2 E85524	probable beta-barr
31	280	4.2	1325	2 A64905	ydek protein - Bsc
32	279.5	4.2	1238	2 A64596	hypothetical prote
33	279	4.2	1369	2 T17504	hypothetical prote
34	277.5	4.2	1487	2 AG2560	hypothetical prote
35	276	4.1	1545	2 B41859	IgA-specific metal
36	273	4.1	2015	2 B81989	hypothetical prote
37	271	4.1	4919	2 T31105	hypothetical prote
38	270.5	4.1	2554	2 A83528	extracellular seri
39	270	4.0	4152	2 T31102	filamentous hemagg
40	269	4.0	1343	2 E90893	hypothetical prote
41	269	4.0	1343	2 D85724	hypothetical prote
42	269	4.0	1536	2 A43855	high-molecular-wei
43	268	4.0	820	2 T17519	cell surface anti
44	267.5	4.0	1190	2 A82615	surface protein Xf
45	267.5	4.0	1910	2 AF0394	probable adhesin h

RESULT 1
B53739
vacuolating cytotoxin precursor - Helicobacter pylori (strain ATCC 49503)
C:Species: Helicobacter pylori
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: B53739; A38137
R:Cover, T.L.; Tumuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.
J. Biol. Chem. 269, 10566-10573, 1994
A:Title: Divergence of genetic sequences for the vacuolating cytotoxin among Helicoba
A:Reference number: A53739; MUID:94193753
A:Accession: B53739
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1287 <COV>
A:Cross-references: GB:U05676; MID:9471727; PIDN:AAAL17657.1; PID:9471729
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Cover, T.L.; Blaser, M.J.
J. Biol. Chem. 267, 10570-10575, 1992
A:Title: Purification and characterization of the vacuolating toxin from Helicobacter
A:Reference number: A38137; MUID:92268100
A:Accession: A38137
A>Status: preliminary
A:Molecule type: protein
A:Residues: 34-56 <CO2>
A:Note: sequence extracted from NCBI backbone (NCBIP:103729)
C:Genetics:
A:Gene: vacA
C:Keywords: cytotoxin
F.1-33/Domain: signal sequence #status predicted <SIG>

Query Match 94.5%; Score 6303.5; DB 2; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;

Qy	1	MEIQQTHRKINRPLVSLALVSLTPQOSHAAFTTIIIPAVGGIATGTAVGTVSGL	60
Db	1	MEIQQTHRKINRPLVSLALVSLTPQOSHAAFTTIIIPAVGGIATGTAVGTVSGL	60

Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120

Qy	121	HYWVGQGNKLEVDMDKAVGTYTSLGLNFTGGDLVNNQKATLRLGQFNQNSFTSYKD	180
Db	121	HYWVGQGNKLEVDMDKAVGTYTSLGLNFTGGDLVNNQKATLRLGQFNQNSFTSYKD	180

Qy 181 SADRTTRVDNFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD 240
Db 181 SADRTTRVDNFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD 240

Qy 241 GATNLASSSVKLMGNVMGRQLQYVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300

Db	241	GATNLNLSNSVKLNGVNWGRLOYGAYLAFYSYSTINTSKVTGEVFNHLTVGDHNAQA	300
Qy	301	GIANKKKTWIGTLDLWQSAGLNIITAPPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS	360
Db	301	GIATSNKKTIGTLDLWQSAGLNIITAPPEGGYKDKPNNTPSQSG-----AKNDKQESS	352
Qy	361	QNNSTQVNPNSAOKTEVOPQTVIDCGPEFAGGKDTVYINIRINTNADGTRIVCGGFKASL	420
Db	353	QNNSTQVNPNSNTQKTEVQTVIDGPFAGGKDTVYINIRINTKADGTTIKVGGFKASL	412
Qy	421	TTNAAHLRIKGGVNLNSQASGRSLIVENLTGNITITVDGFLRVNQVGAYLAGSANSFEF	480
Db	413	TTNAAHLNIGKGGVNLNSQASGRTLVLLENLTGNITITVDGFLRVNQVGAYLAGSANSFEF	472
Qy	481	KAGTDFKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGENTLDFSGVTDKYNINK	540
Db	473	KAGVDFKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGENTLDFSGVTNKVNINK	532
Qy	541	LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	600
Db	533	LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	592
Qy	601	KFKGGEKLVIDFYYSPPNMYFDARINKVETINKLAFQPGQSPNGTSKLMFNNLTLGQNA	660
Db	593	KFKGGEKLVIDFYYSPPNMYFDARINKVETIRKFPASPTPENPGWTSKLMFNNLTLGQNA	652
Qy	661	VMDYSQFSNLTTOGDFINNOGTYINLVRGKGVATLSVGNAAAMFNNDISATGTFYKPLI	720
Db	653	VMDYSQFSNLTTOGDFINNOGTYINLVRGKGVATLWGNAAAMFNNDISATGTFYKPLI	712
Qy	721	KINSAODLIKNTHEVLLKAKIIGYGVSTGTNGISVNLEEQFKERLALYNNNRMDTCV	780
Db	713	KINSAODLIKNTHEVLLKAKIIGYGVSTGTNGISVNLEEQFKERLALYNNNRMDTCV	772
Qy	781	VRNTDDIKACGWAIGDOSVWNPNDYKYLICKANKNIGTSIKTANGSKISVYILGNSTPTE	840
Db	773	VRNTDDIKACGWAIGDOSVWNPNDYKYLICKANKNIGTSIKTANGSKISVYILGNSTPTE	832
Qy	841	NGGNTTNLPNTTNSARNALNAQAPFAOPATPNLVAINQHDFTGIESVFELANRSK	900
Db	833	NGGNTTNLPNTTNARFASYALIKNAPFAH-SATPNLVAINQHDFTGIESVFELANRSK	891
Qy	901	DIDTLVANSAGOGRLQTLTLLIDSHDAGYARKMIDATYSANEITKOLTATTTLNNTIASLE	960
Db	892	DIDTLVANSAGOGRLQTLTLLIDSHDAGYARTMIDATYSANEITKOLTATTTLNNTIASLE	951
Qy	961	HKTSGTLTSLSNAMILNSRLNLSRRHTNIDSFARKLQALKDQKFPASLESAAEVLQYF	1020
Db	952	HKTSSLTSLTSLSNAMILNSRLNLSRRHTNIDSFARKLQALKDQKFPASLESAAEVLQYF	1011
Qy	1021	APKYEPTWVANAIGTSTLNGSNASLYGTSGAVDAYLNGQVFAIVGGFGSGYSGSFNN	1080
Db	1012	APKYEPTWVANAIGGASLNGSNASLYGTSGAVDAYLNGQVFAIVGGFGSGYSGSFNN	1071
Qy	1081	RANLSNGANNTFNGYSRIFANQHEFFEAQAGLSDOSSLNFKSALLDOLNQSYYHLA	1140
Db	1072	QANLSNGANNTFNGYSRIFANQHEFFEAQAGLSDOSSLNFKSALLRDLNQSYYHLA	1131
Qy	1141	YSAATRASGYDFAFFRNALVLPKSVGYSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA	1200
Db	1132	YSAATRASGYDFAFFRNALVLPKSVGYSYNHLGSTNFKSNSTNKVALSNGSSQHLFNA	1191
Qy	1201	SANVEARYYGDTSFYFNMAGVLQEFAPFHVGSNNASLNTFKVNARNPNTHARVMGGE	1260
Db	1192	SANVEARYYGDTSFYFNMAGVLQEFAPFHVGSNNASVLSLNTFKVNATRNPLNTHARVMGGE	1251
Qy	1261	LKLAKEVFLNLGVYVYHLNLSINIGHFASNLGMRYSF	1296
Db	1252	LKLAKEVFLNLGVYVYHLNLSINIGHFASNLGMRYSF	1287

RESULT 2

S44983
vacuolating cytotoxin precursor - Helicobacter pylori (Isolate 185-44)
C:Species: Helicobacter pylori
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S44983; S44102
R:Schmitt, W.; Haas, R.
Mol. Microbiol. 12, 307-319, 1994
A:Title: Genetic analysis of the Helicobacter pylori vacuolating cytotoxin: structure
A:Reference number: S44983; MUID:94335630
A:Accession: S44983
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1291 <SCH>
A:Cross-references: EMBL:Z26883; NID:g472941; PIDN:CAA81528.1; PID:g472942
C:Keywords: cytotoxin
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-1291/Product: vacuolating cytotoxin #status predicted <MAT>

Query Match	93.3%	Score 6222.5	DB 2	Length 1291			
Best Local Similarity	92.7%	Pred. No. 0					
Matches 1202	Conservative 51	Mismatches 38	Indels 5	Gaps 1			
QY 1	METQOTHRKNRPLVSLALV	ALVSLTPQOSHAAFFTVI	IPALVGGIATGTA	VGTVSGL 60			
DB	1	METQOTHRKNRPLVSLALV	ALVSLTPQOSHAAFFTVI	IPALVGGIATGTA	VGTVSGL 60		
QY 61	LSWGLKQAEANKTPDKP	KVRIQAGKGFNEPNKEY	DLYRLLSSKIDCG	DWGNAR 120			
DB	61	LSWGLKQAEANKTPDKP	KVRIQAGKGFNEPNKEY	DLYRLLSSKIDCG	DWGNAR 120		
QY 121	HYWKGQONKLEVDKMD	AVGTYTSLGRNFTGGD	LVNMOKATRLQ	GFNGNSFTSYKD 180			
DB	121	HYWKGQONKLEVDKMD	AVGTYTSLGRNFTGGD	LVNMOKATRLQ	GFNGNSFTSYKD 180		
QY 181	SADRTTRVDENAKNIS	IDNFVEINNRVSGAG	KASSTVLTQA	SEGITSDKNAEISLYD 240			
DB	181	SADRTTRVDENAKNIL	IDNFLEINNRVSGAG	KASSTVLTQA	SEGITSRENAISLYD 240		
QY 241	GATNLASSSVKLMGN	VMMGRLOYGAYLAP	SYSTINTSKVTGE	VNFNHLTVGDKNAAQA 300			
DB	241	GATNLASSSVKLMGN	VMMGRLOYGAYLAP	SYSTINTSKVTGE	VNFNHLTVGDKNAAQA 300		
QY 301	GIANKKTNTGTLDL	MQSAGLNIAPPEGY	KDKPNPTSQS	AKNDKNESAKNDKOESS 360			
DB	301	GIASNKTHIGTLDL	MQSAGLNIAPPEGY	KDKPNKPS-NTTQ	NAKNDRKOESS 355		
QY 361	QNSNTQVINPPNSA	KTEVQPTQVIDGPP	AGGKDTVVN	NRINTNADGTRVGG	KASL 420		
DB	356	QNSNTQVINPPNSA	KTEQPTQVIDGPP	AGGKNTVY	NRINTNADGTRVGG	KASL 415	
QY 421	TTNAHLHIKGGVNL	SNAQSRSLIVENIT	TGNITVDG	PLRVNNOVG	GYALACSSANFEE 480		
DB	416	TTNAHLHIKGGVNL	SNAQSRSLIVENIT	TGNITVDG	PLRVNNOVG	GYALACSSANFEE 475	
QY 481	KAGDTKNGTATFN	NDISLGRFVNLK	VDATNFKIDT	NGCGFNTLDF	SGVTDKVNIN 540		
DB	476	KAGDTKNGTATFN	NDISLGRFVNLK	VDATNFKIDT	NGCGFNTLDF	SGVTKVNIN 535	
QY 541	LITASTNVAYKNF	NINELIVKTINGIS	YGEYTHFSE	DIGSQSRINTVR	LETGTRSLFSGV 600		
DB	536	LITASTNVAYKNF	NINELIVKTINGIS	YGEYTHFSE	DIGSQSRINTVR	LETGTRSIYSGV 595	
QY 601	KFKGGEKLVIDE	YYPWPYFDARN	IKNVEITNK	LAFPGPGS	PWGTSKLMFN	NLTLQNA 660	
DB	596	KFKGGEKLVIDE	YYPWPYFDARN	IKNVEITNK	LAFPGPGS	PWGTSKLMFN	NLTLQNA 655
QY 661	VMDYSQFSLNIT	IOGDFINNOGT	INLYRG	GVKATLSV	GNAAAMFNND	IDSATGFYKPLI 720	
DB	656	VMDYSQFSLNIT	IOGDFVNNOGT	INLYRG	GVATLVN	GNAAAMFFSNND	YSATGFYKPLI 715
QY 721	KINSAQDLIKNTE	HLVLLKAKIIGY	GNVSTGNT	NGINSVNL	BEOPKERL	ALYNNNRMDTCV 780	

Db	716	KINSAODLTKNKEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDICV	775
Qy	781	VRNTDDIKACGMAIGDQSMVNNPDNTKYILGKAWKNIGISKTAGSKISVYILGNSTPTE	840
Db	776	VRNTDDIKACGTAIGNQSMVNNPENKYILEGKAWKNIGISKTAGSKISVHYILGNSTPTE	835
Qy	841	NGGNTTNLTPTNTSNARSANALAQNAAPQAQPSATPNLVAINQHDGTTIESVFELANRSK	900
Db	836	NGGNTTNLTPTNTNKRFSYALIKNAPFARYSATPNLVAINQHDGTTIESVFELANRSN	895
Qy	901	DIDTLYANSAGORDLLQTLTLLDSDHAGYARKMIDATSANEITKQLNTATTTLLNNTASLE	960
Db	896	DIDTLYANSAGORDLLQTLTLLDSDHAGYARKMIDATSANEITKQLNTATTTLLNNTASLE	955
Qy	961	HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLVYQF	1020
Db	956	HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLVYQF	1015
Qy	1021	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN	1080
Db	1016	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN	1075
Db	1081	RANSLNGANNNTNFGYSRIFANQHEHFEAOGALGSDOSSLNFKSALLQDLNQSYHYLA	1140
Db	1076	QANSLNGANNNTNFGYSRIFANQHEHFEAOGALGSDOSSLNFKSALLQDLNQSYHYLA	1135
Qy	1141	YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSSOHLFNA	1200
Db	1136	YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSSOHLFNA	1195
Qy	1201	SANVEARYYYGDTSYFYMAGVQLQEFQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE	1260
Db	1196	SANVEARYYYGDTSYFYMAGVQLQEFQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE	1255
Qy	1261	LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF	1296
Db	1256	LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF	1291
RESULT 3			
G64630			
vacuolating cytotoxin precursor - Helicobacter pylori (strain 26695)			
C:Species: Helicobacter pylori			
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999			
R:Accession: G64630			
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997			
Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.			
Reference number: A64520; MUID:97394467			
A:Accession: G64630			
A:Status: preliminary; nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1290 <TOM>			
A:Cross-references: GB:AE000598; GB:AE000511; NID:g2314019; PIDN:AAD07935.1; PID:g231402			
F.1-33/Domain: signal sequence #status predicted <SIG>			
F.34-1290/Product: vacuolating cytotoxin #status predicted <MAT>			
Query Match 93.2%; Score 6220; DB 2; Length 1290;			
Best Local Similarity 93.1%; Pred. No. 0;			
Matches 1207; Conservative 38; Mismatches 45; Indels 6; Gaps 2;			
Qy	1	MEIQOTHRKNINPLVSLALGVLSITPQOASHAAFTTVIIIPAVGGIATGATGAVTSGV	60
Db	1	MEIQOTHRKNINPLVSLALGVLSITPQOASHAAFTTVIIIPAVGGIATGATGAVTSGV	60
Qy	61	LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNKYDLYRSLSSKIDGGWDMGNAAR	120
Db	61	LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNKYDLYRSLSSKIDGGWDMGNAAT	120

Qy	121	HYWVGQGOQNKLEVDKMDAVGTYTTLGSLRNFRTGGDLVDNNQKATRLRGQFNQNSFTSYKD	180
Db	121	HYWVGQGOQNKLEVDKMDAVGTYTTLGSLRNFRTGGDLVDNNQKATRLRGQFNQNSFTSYKD	180
Qy	181	SADRTTRVDFNAKNISIDNPFVEINNRVGSAGRKASSTVLTLOASEGITSKNAEISLYD	240
Db	181	SADRTTRVDFNAKNISIDNPFVEINNRVGSAGRKASSTVLTLOASEGITSKNAEISLYD	240
Qy	241	GATLNLASSSVKLMGNVWVRGRLOQYVGAYLAPSYSTNTSKVTGEVNFHLLTVGDKNAAQA	300
Db	241	GATLNLASSSVKLMGNVWVRGRLOQYVGAYLAPSYSTNTSKVTGEVNFHLLTVGDKNAAQA	300
Qy	301	GIANKKTTIGTLDLWQSAGLNIAPPEGYKDKPNNTPSQSGAKNDKNESAKNDKQESS	360
Db	301	GIANKKTTIGTLDLWQSAGLNIAPPEGYKDKPNNTPSQSGAKNDKNESAKNDKQESS	355
Qy	361	QNNSTQVINPPNSAOKTEVQPTQVIDGPFAGGKDTVVVNNRINTNADGTIRVGGPKASL	420
Db	356	QNNSTQVINPPNSAOKTEVQPTQVIDGPFAGGKDTVVVNNRINTNADGTIRVGGPKASL	415
Qy	421	TTNAHLHIGKGGVNLNSQASGRSLIVENLTGNITVDGPLRVANNQVGGYALAGSSANFEP	480
Db	416	TTNAHLHIGKGGVNLNSQASGRSLIVENLTGNITVDGPLRVANNQVGGYALAGSSANFEP	475
Qy	481	KAGTDFKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK	540
Db	476	KAGTDFKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK	535
Qy	541	LITASTNVAVKPNINELIVKNGISVGEYTHFSEIDIGSOSRINTVYRLETGTSLSFGGV	600
Db	536	LITASTNVAVKPNINELIVKNGISVGEYTHFSEIDIGSOSRINTVYRLETGTSLSFGGV	595
Qy	601	KFKGGEKLVITDEPYYSPPWNYFDARNIKNVEITNKLAFGPOGSPWGSKLMFNLLTGONA	660
Db	596	KFKGGEKLVITDEPYYSPPWNYFDARNIKNVEITNKLAFGPOGSPWGSKLMFNLLTGONA	655
Qy	661	VMDYSQFSNLTITQGDFFINNOGTINYLVRGKVATLSVGNAAAAMFNNDISATGPKPLI	720
Db	656	VMDYSQFSNLTITQGDFFINNOGTINYLVRGKVATLSVGNAAAAMFNNDISATGPKPLI	715
Qy	721	KINSAODLTKNTEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDTCV	780
Db	716	KINSAODLTKNTEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDTCV	775
Qy	781	VRNTDDIKACGMAIGDQSMVNNPDNTKYILGKAWKNIGISKTAGSKISVYILGNSTPTE	840
Db	776	VRNTDDIKACGTAIGNQSMVNNPENKYILEGKAWKNIGISKTAGSKISVHYILGNSTPTE	835
Qy	841	NGGNTTNLTPTNTSNARSANALAQNAAPQAQPSATPNLVAINQHDGTTIESVFELANRSK	900
Db	836	NGGNTTNLTPTNTSNARSANALAQNAAPQAQPSATPNLVAINQHDGTTIESVFELANRSK	895
Qy	901	DIDTLYANSAGORDLLQTLTLLDSDHAGYARKMIDATSANEITKQLNTATTTLLNNTASLE	960
Db	896	DIDTLYANSAGORDLLQTLTLLDSDHAGYARKMIDATSANEITKQLNTATTTLLNNTASLE	955
Qy	961	HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLVYQF	1020
Db	956	HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLVYQF	1015
Qy	1021	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN	1080
Db	1016	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN	1075
Qy	1081	RANSLNGANNNTNFGYSRIFANQHEHFEAOGALGSDOSSLNFKSALLQDLNQSYHYLA	1140
Db	1076	QANSLNGANNNTNFGYSRIFANQHEHFEAOGALGSDOSSLNFKSALLQDLNQSYHYLA	1135
Qy	1141	YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSSOHLFNA	1200
Db	1136	YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSSOHLFNA	1194
Qy	1201	SANVEARYYYGDTSYFYMAGVQLQEFQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE	1260

Qy 601 KFKGGEKLVIDEEYYPWNTFDARNKTNKBVITNKLAFPGQGSPPWGTSKLMPNNLTLCQNA 660
Db 594 KFKSGEKLVINDEYYPWNTFDARNKTNKBVITNKLAFPGQGSPPWGTSKLMPNNLTLCQNA 653
Qy 661 VMDYSQFSNLTIOGDFINNOGTTNLYLVRGKGVATLSVGNAAAMFNNDDSDATGFKYKPLI 720
Db 654 VMDYSQFSNLTIOGDFINNOGTTNLYLVRGKGVATLSVGNAAAMFNNDDSDATGFKYKPLI 713
Qy 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 714 KINSAQDLIKNTEHVLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 773
Qy 781 VRNTDDIKACGMAIGDQSMYNNPDNYKYLIGKAWKNIGISKTAGNSKISVYYLGNSTPTE 840
Db 774 VRNTDDIKACGMAIGDQSMYNNPDNYKYLIGKAWKNIGISKTAGNSKISVYYLGNSTPTE 833
Qy 841 NGGNTNLPNTTNSARSANALQAONPAQPSATPNLVAINOHDEGTTESVPELANRSK 900
Db 834 NGGNTNLPNTTNNHSAHYALVKNAPFAH-SATPNLVAINOHDEGTTESVPELANRSK 892
Qy 901 DIDPLYANSAGQGRDLLQTLIDSHDAGYARKMDATSAEITKQLTNTATTTLNNTASLE 960
Db 893 DIDPLYTHSAGQGRDLLQTLIDSHDAGYARQMDWTSTGEITKQLNAATDALNNVASLE 952
Qy 961 HKTSGLQTLISLNAIILSLRLVLSRRHTNHDISFAKRLQALKDORFASLESAAEVLQF 1020
Db 953 HKQSGLOTLSLNAIILSLRLVLSRRKHTNHSFAQRQALKGQEFASLESAAEVLQF 1012
Qy 1021 APKYKPTNWMANAIGTSTLUNSGNSASLYTSGAVDAYLNGQVEAIVGFGSGYGSFNN 1080
Db 1013 APKYKPTNWMANAIGGASLNSGNSASLYTSGAVDAFLNGVNEAIVGFGSGYGSFNN 1072
Qy 1081 RANSLNSGANNTPGVVYSRIFANQHEDFEAQALGSDOSSLNFKSALLQDLNQSYYHLYA 1140
Db 1073 QANSLNSGANNANPGVYSREFAQHEDFEAQALGSDOSSLNFKSTLLQDLNQSYYHLYA 1132
Qy 1141 YSAATRASGYDFAFFRNALVLPKSGVSYNHLGSTNFKNSNTNVALKNGSSSQHLFNA 1200
Db 1133 YSATRASGYDFAFFRNALVLPKSGVSYNHLGSTNFKNSNSQSQVALKNGASSQHLFNA 1192
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFAGHVSNNAAASLNTFKVNAARNPLNTHARVMWGE 1260
Db 1193 NANVEARYYYGDTSYFYHLAGVLQEFAGHVSNDVASLNTFKINAAARSPLSTYARAMWGE 1252
Qy 1261 LKLAKEVFLNLGVVYVYHLNLSNIGCHPASNLGMRYSF 1296
Db 1253 LQLAKEVFLNLGVVYVYHLNLSNASHFASNLGMRYSF 1288
RESULT 5
A53739
Hypothetical vacuolating cytotoxin - Helicobacter pylori (strain ATCC 53726) (fragmen
C:Species: Helicobacter pylori
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: A53739
R:Cover, T.L.; Tummuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.
J. Biol. Chem. 269, 10566-10573, 1994
A:Title: Divergence of genetic sequences for the vacuolating cytotoxin among Helicoba
A:Reference number: A53739; MUID:94193753
A:Accession: A53739
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <COV>
A:Cross-references: GB:005677; NID:9471730; PIDN:AAAI7658.1; PID:9471731
A:Note: This strain, designated 87-203, ATCC 53726 tox-, does not possess toxin activ
C:Genetics:
A:Gene: vacA

Query Match 25.4%; Score 1695.5; DB 2; Length 513;
Best Local Similarity 63.3%; Pred. No. 9.3e-80;
Matches 343; Conservative 52; Mismatches 84; Indels 63; Gaps 7;

Db	1696	YPEIKKILGDNFSLSLKGDALNQULTKLTPSDWKNIINELIDNANNVSVQNFNGTL	1755
Qy	166	-----RLGOFNGSFT-----SYKDSADTRTV-----DF	190
Db	1756	IIGATKIGQTDINSAVVFGGLGYQPCDYTDIVCOOKFRETVLGQLLSISADLGJIDTTF	1815
Qy	191	NAKNISIDNFEVNNRVGSGARKA---SSTVLTLOASEGITSOKNAEISLYDGLTN--	245
Db	1816	NAKEIYLTGTGSGNAGWTGGASVTFNSQTSILINQANIVSSQTDGIFSMLGQEGINKV	1875
Qy	246	-----LASSVVKMGV-----HWGRLOYVGYLAYSPSYSTINTSKYTVGEVNF	287
Db	1876	FNOAGIANILGVEAMOSINKAGGLGNLIVTLGSDVYGGYLTPEQKNQTLSQLLGGQNF	1935
Qy	288	NHILTVDGK--NAAQAGIIANKKTNIGTLDLWQS-----AGLNIIAPPGGYKDKPNN	337
Db	1936	DNL-MNDSGLNTAKDLIRQK-----LGFWTGLVGLAGLGIDLQNPKEKLGMSIND	1988
Qy	338	TPSQSGAKNDKNE--SAKNDKQESSONNSNTQVINPPNSAOKTEVQP--TOVIDGPFAGG	393
Db	1989	LLSKKGLFNQITGITSANDIGOVIS---VMLQDIVKPSDALKNDAALGKQMI-GEEL-G	2043
Qy	394	KDTVVNI-----NRLINTNADGCTIRVGGFKASLTNNAHL--HIGKGV-----	434
Db	2044	QDTLNSLESLLQOQTKSVLDKVLAAKLGSIYEOGLGDLIPNLGKGIFFAPYGLSVMQ	2103
Qy	435	--NLSNOAGRSILIVENLTGNITVDGPVRVNNQVGYALAGSSANFEFKAGTDTKNGTAT	492
Db	2104	KGDFSENAQG-VVFVQNSTFSNANGTLSFN-----AGNSLIFAGNNHIAFTNHSGT	2154
Qy	493	FN---NDIS-----LGRFVNLIKVDAAHTANKGIDTNGGENT--LDFSGVTDKVNINKL	541
Db	2155	LNLLSQVSNINVTMLNASGLKINA-TNNNVSVSQGNLFINASCVOQSDPTTASATNPC	2213
Qy	542	ITASTVWVAKNFNIN-----ELIVKTINGISVGEYTHFSEDIGSQSRINTVRLETG	591
Db	2214	TTAQNASSNSASNAPIALNNDESLSVYTAGNF-----NFSGNI-----	2253
Qy	592	TRSLFSGGV-----KFGGGEKLVIDEFYSPWNIFYDARNIKVETNKLAFQPGSPWGTS	647
Db	2254	-----YANGVVDSEKIKG-----SANVKNLYNNAQF-----	2281
Qy	648	KLMFNLTLCQNAVMDYSQ---FSLNLTQGDFFINQGTINTYLVRGKV-----ATLSVG--	698
Db	2282	--QANLNTISNOAVLEKNASFTYNLNTLQGAFF--NNATQKTVEVLQNLVIASNASLSGTIY	2338
Qy	699	-----NAAAMFNNDISATCFYKPLKINSADOLIKNTEHVLKKAKIIGYGNVST	749
Db	2339	GLEVGGALNLLGAIHEN--LENSQTPVPNPLOVGGIINL--NTYTQ-----PFMVSV	2387
Qy	750	GTNGISNVNLEQOFKERLAYNNNNRMDTCVVRNTDDIKACMGAIQDSQVMVNPNDYKYL	809
Db	2388	ANGGTVTLKSSRYID---YN-----INPNLSQSYL	2415
Qy	810	IGRAWNIGTSKTANGSKIS-----VYILGNSTPTEGNGTNTLPTWTSNARSANALA	864
Db	2416	--KLYTLINI---NGNHIEERKNGVTLVYQORVLQDKGLLSVALPNSNNA--SONNLS	2468
Qy	865	QNA-----PPAQPSATPNLVAI--NQHDFGTIESVPELANRSDKIDITLYANS	909
Db	2469	LSVLHNQIKMSYCNKYMDFTPTLDQYIVGIGOSALNQIEAV-----G	2512
Qy	910	GAGORDLQTLIDSHDAG-YARKMIDATSANEI---TKQLNTATYTLNNTASLEHTKSG	965
Db	2513	GNNAIKWLTLMMETKENPLFAPYILEHNSLEILGVTKDQNTFASISJPNPFRNATSL	2572
Qy	966	LOTLISLSNAMILNSRLVNSRRHTNHIDS-FAKRLQALKDOOKFASLESAAEVLQOPAKY	1024
Db	2573	LEMASTQQ---TSRUTKLSDPFRAREGNSERLELKNRFSQD--PNPSEVFVKYSQLS	2628
Qy	1025	EKPTYNWANAIGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGFGSGYSGYSFFNNRANS	1084

```

Query Match      10.1%  Score 674;  DB 2;  Length 2902;
Best Local Similarity 23.6%;  Pred. No. 3.5e-26;
Matches 337;  Conservative 198;  Mismatches 515;  Indels 376;  Gaps 62;

      85  QAGKGNEFPKNEYDLYLSRLSSKIDGWDGN-----AARHYWVG-----GQGN-KL 132
          : : : : :
      1639  QAGSGNNYTKNNNTSESSVSYTNA-----QGNPISALFVYKNGYNFKALGOMALK 1695

      133  EYDMKDVAVGT-YTSLGSLRNFTGGDL-----VNNQKATL 165

```

Db 2629 KHPNMLIQGVGGASFTSGGNTLYGLNVGYDLRVK---SVILGGYVAYGYSGFN--GNI 2693
Qy 1085 LNSGANTNFCVGRIFANQHEFEAOGALGSDQSLNFKSALLQDLNOSYHVLAYSAA 1144
Db 2684 MHSLANVVDVGMVYARAFELKRNEFTLSANEYGGNASHINSNLSLLSVLNQRYNTWTTS 2743
Qy 1145 TRASYGDAFFARNALVKPSVGSYNHLGTFNPKSNSTN---QVALKNGSSSQHLFNA 1200
Db 2744 VNGNYGYDFMKQKSVVLKPOVGLSYHFGLSGMKGMQNPAYQQFVMSHNSPNSVLT 2803
Qy 1201 SANVEARYYYGDTSYFYNNAGVQLQEF-----AHVGSNNAASLNTFKVNAARPLN 1250
Db 2804 NMGLESRYGFKNSYYFVTLRGDLILKAKGDVNVFVGENTL-----LYRKGIFN 2856
Qy 1251 THAVMMGGELKAKEVELNLGVVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1296
Db 2857 TFASVINGGEMHLWRLMYVNAVGLKMLQYQDLYQDLYQDLYQDLYQDLYQDLYQDLY 2902

RESULT 7
B64556
toxin-like outer membrane protein HP0289 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: A64556
R:Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64556; MUID:97394467
A:Accession: A64556
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2893 <TOM>
A:Cross-references: GB:AE000547; GB:AE000511; MID:g2313377; PIDN:AAD07355.1; PID:g231338

Query Match 9.7%; Score 648.5; DB 2; Length 2893;
Best Local Similarity 23.5%; Pred. No. 7.1e-25;
Matches 318; Conservative 189; Mismatches 493; Indels 353; Gaps 57;

Qy 114 DMGNAARHYWYKGGQKLEVDKMDVGTLYTSLGRNFTGGDLVDNMQKATRLGQFNGN 173
Db 1724 DWKN-----INELIDNANNVQFNNGTLIV-----GATQIGQDTN 1761
174 SFT-----SYKDSADRTTRV-----DFNAKNISIDNFEVI 203
Db 1762 SAVVGGLYGQTPCDYTDIVCQKFRGYLGLQLLESSADLGYIDTTNKAKEIYLTGTLS 1821
Qy 204 NNRVSGGAGRKA---SSTVLTLOASEGITSDKNAEISLYDGCATLNLASSSVKL----- 253
Db 1822 GNWGTGGSASVTNSQTSLLNQCANIVSSQTDGIFSMGLGEGINKVFNQAGLANILGEV 1881
Qy 254 -----MGNV---WMGRLOVYVAYLAPSYSTINTSKVTGEVFNHLYVGDK--NAA 298
Db 1882 AVQSIKAGGLGNLIVNTLGSNSVIGGYLTPQKQNTLSQLLQGNQDNL--MNDGLNTA 1940
Qy 299 QAGLIANKNTNIGTLDLWQS-----AGLNIITAPPEGGYKDKPNNTPSOGAKNDKNE 350
Db 1941 IKDLIRQK-----LGFMTGLVGLAGLGGIDQLNPEKLIGSMSINDLLSKGLFNQITG 1994
Qy 351 --SAKNDQESSNNSNTQVNPNSAOKTEVQP--TQVIDGPFAGGKGTVVNI----- 400
Db 1995 FISANDIGQVTS---VMLQDIVKPSNALKNDVYALGQMI-GEFL-QQDILNLSLESLLQN 2049
Qy 401 NRINTNADGTRVGGFKASLTNAAHL--HYGKGV-----NLSNQASGRSL 445
Db 2050 QOIKSVLDKVLAAKGLPIYEOGLGLDIPNLGKGLFAPYGLSQVQKGFSPNAQG--NV 2108
Qy 446 IVENLTGNTIVDGLRVNNOVGGYALAGSSANPEFAKGTDTKNGTATFNNDISLGRFVNL 505

Db 2109 FVQNSTSNANGCTLSFN-----AGNSLIF---AG-----NNHIA---FTN- 2143
Qy 506 KVDAHTANFKGIDTNGGENTLDFSGVTDKVNINKLITASTNVAVKFNFI----- 555
Db 2144 --HAGTQLQLSDQVSNINITLNAS-----NGLKIANNVNSVSGNLFVSCAQOOS 2195
Qy 556 -----NELIV---KTNGISVGEYTHFSDIGSQSRINTVRLETGTRSL-----F 596
Db 2196 DPTTANIANFALSAQSTNGAS-----SNASNAPIALSNNDESLSMAVANDFNF 2245
Qy 597 SGGVKFGGKGLYDIFYSPWNYFDARNIKWBITNKLAFGPOGSPGWGSKLMFNILTL 656
Db 2246 SGNIVANG---VVD---FS--KIKGSANIKNLYLNAQF-----QANNLTI 2284
Qy 657 GQNAVMDYSO---FSLNITIOGDFINNQGTINYLVRGKVATLSVGNAAAMFNNDISAT 713
Db 2285 SNQAVLEKNASFVNNLNIOGAFNN-----ATQKIEVLONLVIANASLST 2331
Qy 714 GYKPLI--KINSADLIKNTHEVLLKAKII---GYGNVSTGTNGISNVNLEEQKPERLA 768
Db 2332 GYGLEVGAGLNSGAIHFNLNTOTPTPLIAEGIIINLNTQTTPFMNVN----- 2381
Qy 769 LYNNNNRMDTCVVRNTDDIKACGMAIGDSMVNPNPNKYKYLIGKAWKNIGISKITANGSKI 828
Db 2382 -----NSMANN--TTYTLKSSRYIDYNINPNLSQSYL 2412
Qy 829 SVYYL---GNSTPTENGNT-----TNLPTNTTSNARSANNAALQNA- 867
Db 2413 NLYTLININGNHIEEKNGALTYLQORVLLQDKGLLSVALP---NSNASONNLSLSVL 2469
Qy 868 -----PFAQPSATPNLVAI--NOHDFGTIESVPFELANRSKIDIDTYANSQAQG 913
Db 2470 YNOVKMSCGKAMDFTPTLQDYIVGIGOGSALNQIEAV-----GGNA 2512
Qy 914 RDLQTLILDS--HDGAYARKMIDATSANEI---TKQLNTATTTLNNTASLEHTSGLOTL 969
Db 2513 IKWLSTLMMETKPPFPFIYLNKHSLEILGVTKDLQNTASLISNPNFRDNTNLLELA 2572
Qy 970 SLSNAMILNLSRLVNLRRHTNHIDSFAKRLQALKDQKFALESAAEVLQYQFAPKYEKPTN 1029
Db 2573 SYTQ---TSRLTKLSDFRSEGESDFSLLE--LANKRFS--DPNPEVYVYQSLSKHPNN 2626
Qy 1030 VWANAIGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNNRANSLNSGA 1089
Db 2627 LHWQGVGASFTSGGNTLYGLNAGYDLRVK---NVLGGVYVAYGYSDFN--GNIMHSLG 2681
Qy 1090 NNTNFGVYSRIFANQHEFEAOGALGSDQSLNFKSALLQDLNOSYHVLAYSAAATRASV 1149
Db 2682 NNVDVGMVYARAFELKRNEFTLSANEYGGNASHINSNLSLLSVLNQRYNTWTTSVNGNY 2741
Qy 1150 GYDPAFFRNALVKPSVGSYNHLGTFNPKSNST--NQVALKNGSSSQHLFNASANVEAR 1207
Db 2742 GYDPMFKQKSVVLKPOVGLSYHFGLSGMKGNDAAAYKQFLMHSNPSNESVLTLMNGLES 2801
Qy 1208 YYGDTSYFYNNAGVQLQEF--FAHVGSNNAASL--NTEFVNAARNPLNTHARVMGGEKLU 1263
Db 2802 KYFGKNSYVFTARLGRDLILKAKGDVNVFVGENTL-----LYRKG--EVFNTFASVITGEMHL 2860
Qy 1264 AKEVELNAGVY 1296
Db 2861 WRLVYVNAVGLKMLQYQDLYQDLYQDLYQDLYQDLYQDLYQDLYQDLYQDLYQDLYQDLY 2893

RESULT 8

B64635

toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64635

R:Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97394467
A:Accession: B64635
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2529 <ROM>
A:Cross-references: GB:AE000602; GB:AE000511; NID:g2314060; PIDN:AAD07969.1; PID:g231406

Query Match 9.6%; Score 642.5; DB 2; Length 2529;
Best Local Similarity 23.0%; Pred. No. 1.2e-24;
Matches 294; Conservative 188; Mismatches 496; Indels 301; Gaps 50;

Qy 149 RNFCTGDLV-----NMOKATLRLGQFNGNSFTS---YKDSAD-----RTT-----186
Db 1421 OTFNGNLIGATQDQVNTNSTSI-WFGNGYSSPCTLSDATCSSFRNTYLQGLLGSTSP 1479
Qy 187 -----RVDFNAKNIIDNFVEINNRVGSAG-----RKASSTVLT-----LQASEGI---228
Db 1480 YLGYINADFKAKSIYITGTIGSGNAFESGGSADVTFSQANNLVLNKANIEAQAOTDNIFNL 1539
229 TSDKNAEISLYDGATLN---LASSSVKL---MGVNMGRLOYVGCAYLAPSYSTINTSKV 281
Db 1540 LQKQIEKIFNOGNLANVLSQVAMEKIKQAGGLGFENALSPLSKELPASLQNETLQOL 1599
Qy 282 TGEVNFNHL--TVGDKNAQAAGIIANKNTNIGTLTLQWSAGLNIIAPPEGYKDKPNNTP 339
Db 1600 IQGNLDDLNNSGVMNAIQ-NIISKLSIFG-----NFV-----TP 1635
Qy 340 S--QSGAKNDKNESAKNDK-----QESSQNSNTQVI-----NPPNSAQKTEVOPT 383
Db 1636 STIENYLAQSKLSMLDDKGLLNFITGGYMNASELSILSVLKDITNPTSLQKDIGVVA 1695
Qy 384 QVIDPFAAGKDTVYNIRNTNADGTRVCGFKASLTNAHLHIGKGVN-LSNQASG 442
Db 1696 NDLLNEFL-GQDVIKKE-----SGLVSNIIINIIISQGLSGVYNQGLG 1739
Qy 443 -----RSLIVENITG-----NITVDGLRVNVNQGVALAGSSANF 478
Db 1740 SVLPSPQLNAKENDGLTLLSPRGLHDFWQGYFNLSNGYVFNSSFSNA-TGSLNF 1798
Qy 479 EPKACTDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTGGNGFNTLDFS-----530
Db 1799 -----VANKSIIFNGD-----NTIDFSKYQCALIF 1823
Qy 531 GVTDKVNLKLTATSNVAVKNFNEILIVKTINGISVG-----EYTHFSEDIGSQSR 582
Db 1824 ASNVSNINITTLNATNGLSLNAGLNNVSVQKGEICVNLANCPTTKNSSSTNSSVTPPTNE 1883
583 INTVLETGTRSLFSGGKFKGGEKL-----VIDEFYSPWNYFDARNIKNVEITNKL 635
Db 1884 SLVVRANFT---FLGATASNGAIDLSQVKNSVIDTLNL---NENAAQLANNLTIN-- 1935
Qy 636 APGPGSPWGTSKLFNNLTIGQNAVMDYSQPSNLTIGDFFINNOGTINY-LVRGKQVAT 694
Db 1936 AFNNASN---STANINGNFTLNQATLS--TNAAGLVNMGNF-NSYGDVLVNLSHSVSHAI 1990
Qy 695 LSVGNAAAMFNNDIDSATGFKYPLIKNSAQDLIKNTEHLLKAKII--GYGNVSTGTN 752
Db 1991 NQAQSATIMANN-----NPLQIOTNKSKEV-GTYTLIDSAKAIYVYNNQITGS 2041
Qy 753 GISNVNLEQFKERIALYNNNNRMDTCVVRNTDDIKACGMAIGDQSMVNNPDNYKYLIG 812
Db 2042 SLDNY-----LKLY-----TLI-----DINGRHVMTDNGL-----2067
Qy 813 AKWNGISKTANGSKISVYLGNSPTPTGGNTTNLTPTNTTSNARNSANNALQAONPAPAP 872
Db 2068 -----TYNGQAVSV-----KDGGLVWGFKD-----SQNGIYVTSILYNKV 2102
Qy 873 SATPNLVAIQHDFCTIESVFELANRSKDIDFLYANSAGQGRDILLQTLIDSHDGYARK 932
Db 2103 KTAIVNDPINNIQAPTLKQYIAQIOGTQGVDSIDQAGGSQAINWLNKIFETKGSPLFAPY 2162

Qy 933 MIDATSAANEITKQLTATTTLNNIASLEHKTSLQTLTSLNSAMILNSRLVNLSSRRHTNI 992
Db 2163 YLESHSTKDLTTIAGDIANTLEVIANPFKNDATNLQINTYTOQMSRLAKLSDSTFPAS 2222
Qy 993 DSFAKRLQALKQDKFA-SLESAAEVLYOPAPKYEPKPTNVWANAIGTSLNNGSNASLYGT 1051
Db 2223 ADHERLEALKRKRFADAIIPNAMDVILKYSQRNRYKNNVWATGVGGASPIINGTGTLYGI 2282
Qy 1052 SAGVDAYLNGQVEAJVGGSGYSYSSFNRRANSLSNGANNNTFGVYSRIFANQHEPDFEA 1111
Db 2283 NVGYDRFING---VIVGGTAAYISGFH---ANITQSSNVNMGVYSRAFIRKSEUTMSL 2337
Qy 1112 QGALGSDQSSALFKSALLQDLNQSHYLAASAATRASGYDFAFFRNALVLKPSGVSYN 1171
Db 2338 NETGYNKTFINSDYPLLSIIINQSYKYDTWTTDKINQYDFEMDKSVIFKPIGLAYY 2397
Qy 1172 HLGSTNFKNST----NOVALKNGSSQHLFNASANVEARYYYGDSYFYMAGVLOEF- 1226
Db 2398 YIGLSCLRGIMDPDIYNQFRANADPNKKSVLTINFALSHRYFNKNSYYFVADVGRDLF 2457
Qy 1227 -----AHVGSNAASLNTFKVNAARNPLNTHARVMMGGELKLAKEVFLNLGVVYLH 1277
Db 2458 INSMGDKMVRFTGNNTLSYRQGR-----YNTFASIITGGEIRLFTFYVNAGIGARF 2510
Qy 1278 NLISNIGHFASNLGMRYSF 1296
Db 2511 GLDYKDNITGNIGMRYAF 2529
RESULT 9
H71879
toxin-like outer membrane protein jhp0856 - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:Date: 12-Feb-1999 sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: H71879
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: H71879
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2399 <ARN>
A:Cross-references: GB:AE001515; GB:AE001439; NID:g4155425; PIDN:AAD06432.1; PID:g415
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0856

Query Match 9.3%; Score 623; DB 2; Length 2399;
Best Local Similarity 22.7%; Pred. No. 1.1e-23;
Matches 296; Conservative 175; Mismatches 479; Indels 356; Gaps 51;

Qy 149 RNFCTGDLV-----NMOKATLRLGQFNGNSFTS---YKDSAD-----RTT-----186
Db 1292 OTFNGNLVIGATQDQVNTNSTSI-WFGNGYSSPCTLSDATCSSFRNTYLQGLLGSTSP 1350
Qy 187 -----RVDFNAKNIIDNFVEINNRVGSAG-----RKASSTVLT 222
Db 1351 YLGYINADFKAKSIYITGTIGSGNAFESGGSADVTFSQANNLVLNKANIEAQAOTDNIFNL 1410
Qy 223 QASEGITSKNAEISLYDGATNLNASSSVKL---MGVNMGRLOYVGCAYLAPSYSTINTS 279
Db 1411 LQOEGII--DKIFNQGNLANVLSQVAMEKIKQAGGLGFENALSPLSKELPASLQDETGL 1468
Qy 280 KVTGEVNFNHL--TVGDKNAQAAGIIANKNTNIGTL-----DLQWSAG- 320
Db 1469 QLIGQNLDLLNNSGVNMEIQ-NIISQKLSIFGFNFTSIENYLAQSLAKSLMLDDKGL 1527
Qy 321 LNTIAPPEGYKDKPNTTPSQSAKNDKNESAKNDKQESSQNSNTQVI-----NPPNSA 375

Db 1528 LNFI-----GGVID-----: : : : : ||| |
Qy 376 OKTEVQPTQVTDGPFAGGKDVVNINRINTNADGTVIRGGFKASLTNTAAHLHGKGGVN 435
Db 1559 QKDTGVVANDLNEFL-GQDVVKLE-----SOGLVSNIIINNVISQGLS 1602
Qy 436 -LSNQASG-----RSLIVENLAG-----NITVDGPLRVNNQVGYYA 470
Db 1603 GYVNOGLGVLPPSLQNALKENDGLTLLSPRGLHDFWQKGYFNFNSGYFVFNSSFSNA 1662
Qy 471 LAGSANPEFKAGTDTKNGTATFNNDISLGRFVNLKVDATANKGIDTNGGNGTDLDFS 530
Db 1663 -TGGSNLF-----VANKSIIFNGD-----NTIDFS 1686
Qy 531 GVTDKVNINKLITASTNVAVKNFENELIVKTINGISVGEYTHFSEDIGSQSRINTVLEET 590
Db 1687 KYQ-----GALIFASNGVS--NINITLNA--TNGLSL-----NAGLNVSVQK 1726
591 G-----TRSL-----FSGVRFKGGKEL--VIDEYY 615
Db 1727 GEICINLANCPTTKNSSPANSVPTNESLSVHANNFTFLGTIISNGAIDLSQVTNNSVI 1786
Qy 616 SPWNYFDARNTK--NVEITNKLATGPOGSPWGTSKLFENNLTLGQNAVMDYSQFSNLTIQ 673
Db 1787 GTLLNENATLQANNLITIN--AFENNASN--STANIDGNFTLQQAATLS--TNASGLNVM 1840
Qy 674 GDFINNOCTINY-LVRGKQVATLSVGNAAAMFENNDDISATGFYKPLIKINSADLLKNT 732
Db 1841 GNF-NSVGDVFNLSHVSIAIITQGTATIMANN-----PLIQFNASKEV--GT 1889
Qy 733 EHVLLKAKII--GYGNSTGTNGISNVLBEQFKERLALYNNNRMDTCVVRNTDDIKAC 790
Db 1890 YTLIDSAKAIYGYNNQITGSSLDNY-----LKL-----ALIDNGK 1928
Qy 791 GMATGDOSWNNPNKYLIKAKNKGISKTANGSKISVYYLGNSTPTENGNTNLPT 850
Db 1929 HVMVTNGL-----TYNGQAVS-----KDGGLVWGFKD 1957
Qy 851 NT-----TSNARSANALAOAPFAQSPATPNLVAIQHDFTGIESFELANRSKDIDL 905
Db 1958 SONQYIITSILYKVKTAVSNDPINNQAP-----TLKQYIAQGVQSVDSI 2005
Qy 906 YANSAGRODLLQTLIDSHDAGYARKMIDATSANETKQLNTATTTLNNTASLEHKTSG 965
Db 2006 DQAGNOAINLNLKIFETKGPLFAPYPLESHSTKLTITAGDIANTLEVIANPNFKDA 2065
Qy 966 LQTLSLNAMTLNRLNLSRRHTNHIDSFAKRLQALKDKQEA-SLESAAEVLQFAPKY 1024
Db 2066 TNLQINTYTQMSRLAKLSDTSTFARSDFLERLEALKNKRFAADAI PNAMDVILKYSORN 2125
Qy 1025 EKPNTVWANAIGGTSLNNGSNASLYGTSGVDAYLNGOVEAIVGFGSGYSGYSFNNRANS 1084
Db 2126 RVKNVWATGVGGASFTSGGTGLYINGVYDRFIK-----VIVGYYAAYGSGFH--ANI 2180
Qy 1085 LNSGANTNFVGYGRIFANQHEFDEAQAALGSDQSSLNFKSALLQDLNOSHYLAYSAA 1144
Db 2181 TQSGSSNVNVGVYRAFIKRSELMTSLNETWGNKTFINSYDPLLSIINQSYRYDTWTD 2240
Qy 1145 TRASYGVDFAFRNALVLKPSGVSYNHLGSTNFKSNST-----NOVALKNQSSSQHLENA 1200
Db 2241 AKINYGDFMFKDSVIFKPOVGLSYIYIGLUGLGRIMDDPIYNOFRANADPNKKSVLTI 2300
Qy 1201 SANVEARYYYGDTSYFVNAGVLQEF-----AHVGSNNAASLNTFKVNAARNPLN 1250
Db 2301 NFALESRYHFNKNSYFYVIADVGRLDFINSMDKWKVRFIGNNTLSYRGGH-----YN 2353
Qy 1251 THARVMGGEKLAKAEVFLNGLVGYVYLNHLNLSIGNHGFASNLGMYYSF 1296
Db 2354 TFSAITGGEIRLFKTFVNAGIGAREGLDYKDINITGNIGMRYAF 2399

B64596
toxlin-like outer membrane protein HP0610 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64596
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: B64596
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1943 <TOM>
A:Cross-references: GB:AE000575; GB:AE000511; NID:g2313730; PIDN:AAD07675.1; PID:g231
C:Genetics:
A:Start codon: GTG

Query Match 8.7% Score 583 DB 2 Length 1943;
Best Local Similarity 22.3% Pred. No. 9.2e-22;
Matches 330; Conservative 212; Mismatches 540; Indels 398; Gaps 70;
Qy 6 THRKINRPLVSLALVGALVSITPQO-SHAFFTTVIIPAIVGGIATG-----TAVGT 56
Db 673 TYNQAQNPQALHIYNQ--AITKODLNMIASLGKEFLPKIANLLSSGALDNLNPSNPSET 730
Qy 57 VSGLL-SWGLKQAEANKTPDKPKVRIQAGKGFNEFPN-KEYDLYRSLSSKIDGWD 114
Db 731 LFGIFERYGITLNGEN-----WK-SLLKIIINFSTNTNYDFSQNL---VWGA-- 774
Qy 115 WGNARHWYKGGQONKLEV-----DMKD--AVGTTVLSLRNFTGGDLVDVNMOKATLR 166
Db 775 -----IKEGDTNKSVVMFGGEGYKEPCAVGDNTQCMFROTNLQGL---LHSSTPY 822
Qy 167 LGQFNGNSFTSYKDSADRTTRVDNFNAKNISIDNEFINNRVSGSAGRKAS-----STVLT 221
Db 823 LGINAN-----FRAKNIYITGTIGSGNANGSGGSANVSFESGTLNVLN 866
Qy 222 LQASEGITSOK-----NA-----EISLYDGA-TLNLAASSVKLM 254
Db 867 QAKIDAGOTDKIFSLYLGQGGIEKLFEGKGLGNALNSIIEESLNDNAIPKOLANMIPKDF 926
Qy 255 GNVWNGRLQYVGAYLAPSYSTINTSKVTGEVNFNHLTVGD-KNAQAAGIIANKKTNIQTL 313
Db 927 GSKTSSL-----LSPT-----EYN-NLLGVSAFKNAIME--ILNSKT-VG-- 963
Qy 314 DLWQSAG-LNIIAPPE-----GGYK-----DKPNNTPSOSGAKND 347
Db 964 DVFGENGLLNALDPTERRKIDOMLLEQIOAHSSGFEKFIKTLGTIENVENFINNWKQOS 1023
Qy 348 KNESAKN-----DKQESSNNSTQVINPPNSAKQT---EVQPTQVIDGP----- 389
Db 1024 LSSFANNFVPGGLNQALDKIGSSSDAKDLQ-----NFLDKTTFGDILNQMTQEPALINKL 1078
Qy 390 --FAGGQDVTVINR-RINTNADGTVIRGGFKASLTNTAAHLHGKGGVN--LSNOASGR- 443
Db 1079 ISWLGPDQLSVLVNIALNSITNPSKELTSTISSIGERKALNDLLGDGVVKNKTMNQVLQOM 1138
Qy 444 --SLIVENLTGNITVDG-----PLRVNNQ-----VGGYALAGS-----SANFEKAGTD 485
Db 1139 INKTIADKFGVGYOQGLSLQSLQDELKKGMLGSLGSLGRGLNHLWQRNFPNFAVDY 1198
Qy 486 TKNTATFNNDISLGRFVNLKVDATANKGIDT-----GNGFNTLDSGYT-DKVN 537
Db 1199 LFTNNSFSFN--ATGGELNF-VAKGSIIFNGKNTINTFTQVQGLSFISKDFSLDLOTLN 1255
Qy 538 INKLITASTNVAVKNFENELIVKTINGISVGEYTHFSEDIGSQSRINTVLEET---GTR 593
Db 1256 -----ATNGLTLNAPKNDISVQKQICVNVNLCMGEKKAHSSSATAPTNETLEANANN 1308

QY 594 SLFSGGVKFKGKELVIDEYFYSPWNYFDARNIKNVEITIKLAFPGQSPWGTSLMFN- 652
Db 1309 FAFLGAIRANG---LVD-----FSKVQNTTI-GTLDLGNAT-FKANHLIVNN 1352
QY 653 -----NLTGONAVMDYSQFSNLTIOGDF-----INNQTINVLVRG 689
Db 1353 AFNNNSNYRADISGNLNVKGAALSTNE-NGLVGDFKSEGLSIFNLNKNKNOTIINVA 1411
QY 690 GKVATLSYGNAAAMFNNDIDSAT-----GFYKPLIKINSADOLIK-----N 731
Db 1412 GNSTTMSYNNQALIHFNQLKOGATLINAKRMILGYDNOIIRGSLDYLKLYLIDPN 1471
QY 732 TEHVLLKAKIIGYGVSTGTNGISNVNLEEOPEKRELALYNNNNMDTCVVRNTDDIKACG 791
Db 1472 GRMQLNGDSLSDYDQPV---NIKDGLVVSFKD-----NOCOMVYSIL--YDKVQ--- 1518
QY 792 MAIGDQSVNPNPDNYKYLIGRAWKIGISKTANGSKISVYVYLGNTPTPENGNTNLPN 851
Db 1519 VSVSDKPMIDIHAPSLEYIKYIQGSAGLDAIKSAGNNSILWL-NELFVAKGGNPLFAPY- 1576
QY 852 TTSNARSANNAQAONAPAPQSPATPNLVAIQHDFGTIESPELANRSKDIDTYANSGA 911
Db 1577 -----YLDNPTTHIVTL-----MKDI-----TSA 1596
QY 912 QGROLLOTLIDSDHAGYARKMIDATSANETKQNTATTTLNNTIASLEHKTSGLOLTL 971
Db 1597 LGMLSKPNLKNNSTDA-----LQNTVYTOQMSRLAKLSNFASFST--- 1637
QY 972 SNAMTLNRLNLSRRHTNHDISFAKRLQALKDQKFA-SLESAAEVLYQFAPKYEKPTNV 1030
Db 1638 -----DFSERLSSLNQRFADAIAPNADVILKYSORDKLKNL 1675
QY 1031 WANAIGGTSLNNGSNASLYGTSGAVDAYLNGOVEAIVGFGSYGYSSFNRRANSLSNGAN 1090
Db 1676 WATGVGGVSYFVENGITLYGVNVGYDRKIG---VIVGYAAYGYGFERIT--NSKSD 1730
QY 1091 NTNFVGVYRIFRANQHEFDEAAGALGSDQSSLNFKSALLQDLNQSYHLYLAYSAAFRASYG 1150
Db 1731 NVVDGLYARAFIKKSELTFSVNETWGANKNQISSNDTLLSMINQSYKYSTWTTNAKVNYG 1790
QY 1151 YDAFFRNLVLKPSGVSYNHLGTSNPKS---NST-NOVALKNGSSSOHLENASANVEA 1206
Db 1791 YDFMKNKSIIILKPOIGURYIYIGTLEGVMHNALYNOFKANADPSKKSVLITIELALEN 1850
QY 1207 RYVYGDTSFYFYNMAGVLOEFARHVS-----NNAASLNTKVNAAARNPLNHARVM 1256
Db 1851 RHYFNTSYFYAIGGFGROLL-VNSMGDKLVRFIGNNTLSYKKGEL-----YNTFASIT 1903
QY 1257 MGGELKLAKEVFLNLGVVYLHNLISNIGHFASNLGMYSF 1296
Db 1904 TGGEVRLFKSFYANAGVGARGLDYKMINITGNIGMRLAF 1943
RESULT 11
D71917
toxin-like outer membrane protein jhp0556 - Helicobacter pylori (strain J99)
A:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: D71917
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: D71917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3194 <ARN>
A:Cross-references: GB:AE001488; GB:AE001439; NID:g4155100; PIDN:AAD06134.1; PID:g415510
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0556

Query Match 8.7%; Score 580.5; DB 2; Length 3194;
Best Local Similarity 21.9%; Pred. No. 2.6e-21;
Matches 328; Conservative 208; Mismatches 543; Indels 419; Gaps 65;
QY 2 EIQOQTHRKINRPLVSLALVGLVSIPTQOSHAAFFTVIIPAIVGGIATGTAVGTVSGLL 61
Db 1913 EIPGIYNAQNQPLQALHIYNAIS-KQDLNMIASLKGKFLPKVAKLIASG----- 1962
QY 62 SWGLKQAEAN-KTPDKPKVWRIOAGKGFNEFPNKEYDL-----YRSLLSKIDGGWD 114
Db 1963 -----LDNLNLSPDFETIFSIL-----KEYGITLNOANWKSLL-KIIN--N 2002
QY 115 WGNAAARHYW-----VKGGOONKLEV-----DMKD--AVGTVTSLGRNFTGDDLD 157
Db 2003 FSNTANYHESQSLVVAIKEGTNTNSVWFGGDKYKNPCAVGDMTCOMFQTNUGQL- 2061
QY 158 VNMQKATLRLGNGNSFTSKDSADRTTRVDFNNAKINISIDNFVEINNRVSGGAGRKA-- 215
Db 2062 --LNSSVPLYGINAN-----FKAKNIYITGTIGSGNAGSGGSANVSF 2103
QY 216 -SSTVLTLQ-----ASEGI-----TSKNAEISLYDGA-TLN 245
Db 2104 ESATNLVLNQAANIDAGTDKIFSYLCKEGIDKLFGEKGLGNVLSNIVYBESLNDNAIPKD 2163
QY 246 LASSSVKLMGNVMGRLOVVGAYLAPSYSTINTSKVTGEVN-----FNHL 290
Db 2164 LANMPKDLGSKTSSL-----LSPT-----EVNLLGVSAFKNAIMEILNSK 2206
QY 291 TVGD---KNAAGAIANKKTNIGTLDLWQ-----SAGLNIIAPPEGYKDKPNTPSQSG 343
Db 2207 TVGDVFGENLLNALDPVKRKEIDQMLLQIAHSSGFEKIVKTGIGIENVENFINNWYG 2266
QY 344 AKN-----DKNESAKNDKQESSQNSNT-----QVNNPP----- 372
Db 2267 KOSLSFANNFPVGLNQALDKIGSSDAKDLQSFDKTTFGDIILNQMINAPLKNLIS 2326
QY 373 -----NSAOKTEVQTPQVIDGPFAGGKDTVNV-----INRNTN-----AD 408
Db 2327 WLGPQDLSVLVIALNSITNPSSKELLGALSGMGQKVLNDLLGEGVVNKTMSNOVLQOMIN 2386
QY 409 GTRVGGF-----KASLTTNAAHLHIGKGVNLSNQASGRSLIYENLTGNITVDGCLR 461
Db 2387 KTIADKFGVGVYHQGLGSLPKSLQDELAKLGMGSLLPKGLHNLWQKGFNFVAKNHVF 2446
QY 462 VNNQVGGYALACSSANFEKAGTD-TKNGTATFNNDISLGR-----PVNLKVDATHTA- 512
Db 2447 VNNSLFSNATGG---ELNFAVGKSIIFNGKNTINFTQYQGRLSFVSKDFSNISLDTLNAT 2503
QY 513 -----NFKGIDTNGNGFNLTLD--FSGVTDKVINKLKITASTN---VAVKNFNIN 556
Db 2504 NGLTINASKNDISVOKGQICVNVLDGMTAKGKTQTNSSSSATAPTNETLEVSANNAFEL 2563
QY 557 ELIVKTNGISVGEYTHFSDIGSQSRINTVRLETGTRSLFSGGVKFKGKELVIDEYFYS 616
Db 2564 GFI-RANGL-----VDFSK-VLQNTTIGT--LDLGNATFK-----ANNLIVNAFNN 2607
QY 617 PNWYFDARNIK-NVEI-----TNKLAFCGPOGSPWGTSKLMPN-----NLTGQ 658
Db 2608 NSNY--RANISFNFNVAKGATFSTENGLNVGNFNSEGLFIFNLNPNTHQTIINYT-GT 2664
QY 659 NAVMDYSQFSNLTIOGDFTNNGTINYLVRGKGVATLSVGNAAAMFNNDIDSATGYKP 718
Db 2665 STMSY-----NNQALINFNTQ-LKQAYTILINNRWVGYQNDOTILG----- 2706
QY 719 LIKINSADOLIK-----NTEHVLLKAKIIGYGVNVTGTNGISNVNLEEQFKERLALYN 771
Db 2707 ---GSLSDYLKLYLIDFNGRMQLNGDSLSDYDQPV--SIKDGGLVVSFKD-----N 2754
QY 772 NNNRMDTCVVRNTDDIKACGMAIGDQSVNPNPDNYKYLIGKAWKNIGISKTANGSKISYV 831
Db 2755 QGQWVYSSIL--YDKIQ---VTVSDKPMISQAPSLSEYVVKRIQGSAGLNAIKSAGNNSIM 2809

Qy	832	YLGSTPTPENGNTNLPTNTTNSARSANNAQAQPAQSPATPNLVAINQHDGFTIES	891
Db	2810	WLSELFPAK--GGNPLFAPY-----YLQDNPTHEIVTL-----	2840
Qy	892	VFELANSKIDIT---LYANSAGQR--DLQTLILSDHAGYARKMIDATSEITKQL	946
Db	2841	-----MKDITTSALGMLNSNLKNNSTDVL-----	2866
Qy	947	NTATTNNIATSLSEHTSGTLTSLSNAMILNSRLVNLRRRTHNIDSFAKRLQALKDOK	1006
Db	2867	NTYQQMSRLAKLSNPFSDST-----DFSERLSSLNQKR	2901
Qy	1007	FA-SLESAAEVLQYAPKYEKPTNVMANAIGCTSLNNGSNASLYGTSAGVDAYLNGVEA	1065
Db	2902	FADAVPNAMDVILKYSDORKLNNLWATGCGSPVENGTGTLGCVNVGYDREVRG--V	2958
Qy	1066	IVGFGSGYGYSSNNRANSLSNCANNTPGYYSRIFANQHFEDFEAQOALGSDOSSLNFK	1125
Db	2959	IVGYYAAYGSGFYERITS--SKSDNVDVGLYARAFIKKSELTFSVNETWGANKTQISSN	3016
Qy	1126	SALLQDLNQSHVLYLAYSAASTRASGYDFAFFRNALVLKPSVGVSNHLGSTNFKS--NST	1183
Db	3017	DALLSMINQSYKYSTWTTTRAKVNYGYDFMFKKKSIIILPQIGLRYYYIGMSGLEGVNNV	3076
Qy	1184	--NOVALKNGSSQHLFNASANVEARYYGDTSFYFMYNAGVLQBEFA--HYGSSNNAASLNT	1239
Db	3077	LYNQFKANADPSKKSVLTIDFALENHRHYNTNSFYAIGGVGRDLLYNSMGDKLVRFIGN	3136
Qy	1240	FKVNAARNPL-NTHARVMGGELKLAKAEVFLNGLGVVYVLIHNLISNIGHFASMLGMYSF	1296
Db	3137	NTLSYRRGDLNTFANITTTGGEVRLFKSPYANAGYARGFLDYKMDIIGNIGMRLAF	3194
RESULT 12			
E97835			
Hypothetical protein rompB [Imported] - Rickettsia conorii (strain Malish 7			
C:Species: Rickettsia conorii			
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001			
C:Accession: E97835			
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.;			
Science 293, 2093-2098, 2001			
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prow			
A:Reference number: A97700; PMID:21442074; PMID:11557893			
A:Accession: E97835			
A:Status: preliminary			
Molecule type: DNA			
Residues: 1-1655 <R>			
Cross-references: GB:AE006914; PIDN:AAL03623.1; PID:gl5620207; GSPDB:GN00			
C:Genetics:			
C:Gene: rompB			

Query Match	5.0%;	Score 335;	DB 2;	Length 1655;
Best Local Similarity	19.5%;	Pred. No. 4e-09;		
Matches 274;	Conservative 177;	Mismatches 531;	Indels 420;	Gaps 59;
Qy	146	SGLRNFTGGDLDVNMOK-ATLRLGOFNGNSFTSYKDSADRTTRVDFAKNTSIDNFVEIN	204	
Db	211	NGTLVNTNGFTKYSSKSFATVNVINIGDQIGIMFNTDADNVNTLNLQANGATI----	266	
Qy	205	NRVSG----AGRKASSTVLTLQAS-----EGI-----TSDKNAEISLYDGA-----TLN	245	
Db	267	GTDTGTGRVLVLLSKNAAATDFNVGTSLGGLNKGIIIEFNFAVAVNGQLKAGANAAVIGTNN	326	
Qy	246	LASSSVKLMGNVMGRLOYGAYLAPSYSTNTSKVTGEVFNHL-----TVGDKNAA	298	
Db	327	GAGRAAGFVWVDNGKVFATIDQGVYAKDMVLOSANAVGQVFNFRHVLVDVGTGTTFATKFA	386	
Qy	299	QAGIIANKTKNTIGTLDLQWSAGLNIIAPPE-----GGYKDKPNNTP-----SQS	342	
Db	387	-SKVAITQNSNFGTDFDGNLAA-QIIVENTMTLNGNFTGDASNPCGTAGVITTFDANGTLA	444	

Qy	343	GAKNDKNESAKNDKQSSONNSNTQVINPPNSAQ-----KTEVQPTQ	384
Db	445	SASADANVAVTNITATEASGAGVQVQLSGTHAAELRLGNAGSVFKLADCTGVINGKVNQTA	504
Qy	385	VIDGPFAGGKDTV---VNTNRINTNADGRTIRVGGF-----KASLTNTTAAHLHIGKGVN	435
Db	505	LVGALAAGTITLDGSATITGDIGNAGGAAALOGITLANDATKTLTLLGGANI--IGANGGT	563
Qy	436	LSNOAGSRSLIVENLTGNITVDCPLRVNNQVGYALAGSSANFER-----KAGDTKNG-	489
Db	564	INFOANGGTIKLSTQNNIVVDFDLAIAIDQGTGVVDASSLTNAQTLTINGIKIGTVGANNK	623
Qy	490	-----TATFNNDISLGRV-----	508
Db	624	TLQGFNIGSSKTVLSDGDVAINELVINGNGAVQFAHNTYLITRTTNAAGCKIIFNPVN	683
Qy	509	AHTANFXGIDTNG-----GFNTLDFSGVTDKYNINKLITA-STNVAVNFNINELIV	560
Db	684	NNTTLATGTLNLSGATNPLAEINFGSKGAANVDTVLNVGKVNLYATNITTTDANVGSFIF	743
Qy	561	KTNGISGEVTHEFSEDIGSO--SRINTVLETGTRSLFSGGVKFKGGEKLVIDEFYFSPW	618
Db	744	NAGGTNI-----VSGTVGGQGNKFTVALDNGTTVKFLGNATFNGNTTIAANSTLQIGG	798
Qy	619	NY-----FDARIKNVEITN---KLAFGPGQSP-----WGTSKLFMNNLTGQ	658
Db	799	NYTADFVASADGTGI--VEFVNTGPTITVLNKOAAVPNALKQITVSGPGNVVINEIGNAG	856
Qy	659	N---AVMDYQSFNSLTIOGDFINNOGTINYLVRGKQVATLSVGNNAAMWFNNDI--DSAT	713
Db	857	NYHGAVTDTTAFENSSL-----GAVVFLPRG--IPFDAGNRIPIUTIKSTVGKNTAT	906
Qy	714	GFYKPLKIKNSAODLKTNEHVLLKAKIIGY-----NVSTGTNGISNVNLEE---	761
Db	907	GFDPVSIVILGVDSVADGQVIGDQNNIVLGLGSDNLIIVNATTLYAGITINNNGTV	966
Qy	762	-----QFKERLAL--YNN-----NNRMDTCVVRNTDDIKA	789
Db	967	TLSGGIPNTPGTVGLGTGIGAKPKQVFTTDDYNNLGNIIATNATINDGTVTTTGGIAG	1026
Qy	790	CG-----MAIGDQSMVNNPNKY---LIGKAKNKIGISKTANGSKISVYVLGNS-----	836
Db	1027	IGFDGKITLGS---VNGNGNVRFDVGDILSHSTSMIGTTKANNG--TVTYLGNFAVGNIG	1080
Qy	837	-----TPTEN-----GGNTT-----NLPT	850
Db	1081	DSDTPVASVRFSGDGAGLQGNISQVDFGTYNLIGNSNVILGGTTAINCKINLRT	1140
Qy	851	NTTSNARSANNALAQNAPFAPSATPNLVAINQHDFTGIESYFELANS-----	899
Db	1141	NLTIFA-SGTSTWGNLT-----SIETTTLTLANIGNIVILEGAQ	1179
Qy	900	-----KQIDTLYAN-SSAQGRDLLQ-----TLLIDSHDAGYARKMID	935
Db	1180	VNATTTGTTTIIKVQDNANANFSGTQTYTLTLOGAREFNGTLGPNFVVTGSRNRFVNYGLIR	1239
Qy	936	ATSANEITKOLNTA-----TTLLNNTIASLEHKTSGLOTLSLS	972
Db	1240	AANDGYVITNTNAENVVNDIANSEFGAPGVQNVTFVFNATNTAAYNLLKAKNSAN	1299
Qy	973	NAMLNRLVNLRRHITNHDISFAKRLQAKDQKFASL-----	1010
Db	1300	SANFVGAIVTDTSAATNAQLDVAKDIOAQLGNRLGALRYLGTPTETAEMAGPEAGIAPAA	1359
Qy	1011	-----ESAAEVLVQFAPKYEKP--TNWANAICGTSFLNNGSNASLYGTSAGVDAYLNGQV	1063
Db	1360	VAAGDEAVDNVAYGI---WAKPPYTDHAQSKKGGLA---GYKAKTGTGVVIGLDTLANDNL	1413
Qy	1064	EAIVGFGSGYSGYSPFNRRANSLSNG-----ANNTNEGVYSRIFANOHEFDPEAQ--ALGSD	1118
Db	1414	--MIG--AALIGITKTDIKHODYKKGDKTVDVNGFSLYLG---AOLVKNFPAQGSALFSL	1466
Qy	1119	QSSLNFKSALLQDLNOSY-----HYLAYSAATRASYCYDFAFERNALVLKPSGVGSYN	1171

Db 1467 NOVKNRSQRYFFDANGNNKSKQIAAGHYDNMTFGNLTGVYDYNAMQGVLLV-TPMAGLSYL 1525
QY 1172 HLGSTNEKSNSTNOVALKNGSS 1193
Db 1526 KSSDENYKETGTT-VANKOVNS 1546

RESULT 13
S07575
outer membrane protein B - Rickettsia rickettsii
N:Alternate names: 120K surface-exposed-protein
N:Contains: 32K beta peptide
C:Species: Rickettsia rickettsii
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Dec-1998
C:Accession: S07575; S18579; A43869
R:Gilmore Jr., R.D.; Joste, N.; McDonald, G.A.
Mol. Microbiol. 3, 1579-1586, 1989
A:Title: Cloning, expression and sequence analysis of the gene encoding the 120kD surface
A:Reference number: S07575; MUID:90136087
A:Accession: S07575

Molecule type: DNA
Residues: 1-994, 'DLKLEH', 1001, 'GS', 1005-1300 <GIL>
Cross-references: EMBL:X16353
A:Note: this sequence has been revised in reference S18579
R:Gilmore, R.D.; Joste, N.; McDonald, G.A.
Mol. Microbiol. 5, 3089, 1991
A:Reference number: S18579; MUID:92236427
A:Contents: erratum
A:Accession: S18579
A:Molecule type: mRNA
A:Residues: 994-1004 <GIL2>
R:Hackstadt, T.; Messer, R.; Cieplak, W.; Peacock, M.G.
Infect. Immun. 60, 159-165, 1992
A:Title: Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein
A:Reference number: A43869; MUID:92104668
A:Accession: A43869
A:Molecule type: protein
A:Residues: 1008-1027 <HAC>
A:Note: sequence extracted from NCBI backbone (NCBIP:74388) and corrected to correspond
C:Keywords: glycoprotein
F:2-1007/Product: outer membrane protein B #status predicted <MAYO>
F:1008-1300/Product: 32K beta peptide #status experimental <MATB>

Query Match 4.7%; Score 314.5; DB 2: Length 1300;
Best Local Similarity 21.0%; Pred. No. 3.2e-08;
Matches 274; Conservative 164; Mismatches 489; Indels 377; Gaps 60;

QY 111 GWDWGNARHYVWVGGOONKLEVDKDAVGYTTLISGLRNFTEGDLVDNMQRATRLRGQF 170
44 GNTDFGNLA-----AQIKVPNAI---TLTG--NFTG---DASNPGNTAGVITF 83

QY 171 NGNSFTSYKDSADRTVRDFNAKNTSIDNFVNNRVSGAGRKASSTVLTLOASEGITS 230
Db 84 DANG-TLESAD-----ANVAVTNITAIASGAGVYQLSGT-----HAAELRLG 128

QY 231 DKNAEISLYDGATLNLASSVKVLMGVNMRQLQYQYGVAYLAPYSTINTSKVTGEVNFNHL 290
Db 129 NAGSIFKLDAGVINGKVNQTLVG-----GALAAGTITLDGSATITGDIG---- 174

QY 291 TVGDKNAQAAGTIANKNTNIGTLDLWQAGLNIIAPPEGGYKDKPNTPSQSGAKNDKNE 350
Db 175 NAGGAALQRITLANDAKKTLT-----GGANII-----GAGGGTIDLOA 214

QY 351 SAKNDKQESSQNNSTQVINPNNSAQKTEVQPTQVIDGPFAGKDTQVNVININTNAOGT 410
Db 215 NGGTIKLTSTQNNI-----VVDFDLA-----IATDQTCV 243

QY 411 IRVGGFKSLTNAHLHI-GK-GGVNLSNQASGRSLIVENTGNITVDGPLRVNNOVGG 468
Db 244 V-----DASSLNTAOTLTNGIKGTIGANNKTLGOFNIGSSKT--VLSNGNVAINELVIG 296

QY 469 YALAGSSANFEFKAGDTGKTNGTA-----TFNNDISLGRFVNKLKVDARTAFKGDITG 520
Db 297 NDGAVQFAHDTYLI-TRTNAAAGQGLIFNPVYNNGTTLAAGTNL--GSATNPLAEINFG 353

QY 521 NCGFNTLDFSGVTDKVNINKLITASTNVAVKNFNINELIVKTNISVGEYTFHSEIDIGSQ 580
Db 354 SKGVNVDTVLVNVEGVNL-----YATNITTTDANVGSFVNAGGTNI-----VSGTVGGQ 403

QY 581 --SRINTVRLGTGTRSLFSGGVKFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAF- 637
Db 404 QGNKFTVALENGTTVKFLGNATFNGNTTIAANSTLQICGNY-TADCVASADGTGIVEFV 462

QY 638 --GP-----QGPMGTSKLMFNNTLQON---AVMDYQSFQSNLTIQGDF 676
Db 463 NTGPTTVTLNKEAAPVNALKQITVSGPGNVVINEIGNAGNHGAVTDTTAFENSSL---- 518

QY 677 INNOGTINYLVRGGKAVATLSVGNAAAMFNNDIDSAT--GFYKPLIKINSADOLIKNTEH 734
Db 519 ----GAVVFLPRG--IPENDAGNTMPLTIKSTVGKNTAKGFDVPSVVLGVDSVIADGQV 572

QY 735 VLLKAKIIGY-----VYSTGTNGISNVNLEB----- 761
Db 573 IVDQNNIVGLGSDNGIIVNATTLYAGISTLNNOGTVTLSGGVPNTPGTVYGLGTGIG 632

QY 762 --QFKERLAL--YNN-----NNRMDTCVVRNTDDIKACG---MAIGDQSMVNNPDNY 806
Db 633 ASKEKQVTFTTDYNLGNLIATNATINDGVTVTGTGGIAGIGFDGKITLGS---VNGNGNV 689

QY 807 KY---LIGKAWKNIGISKTAGSKISVYVLGNSTPTENGCGNTNLP---TWTTSNARSA 859
Db 690 RFADGILSNSTSMIGTTKANNG---TVTYLGNAFVGNIGSDTPTVASVRFTGSDSGAGLQ 746

QY 860 NNALQAAPFAQPSATPNLVAIHOHF-----GT----- 888
Db 747 GNIYSQVIDF---GTYNLGIIVNSNIIILGGGTATNGKIDLVNTLTTFASGTSTWGNNTS 802

QY 889 IESVFELANRS-----KIDITLYAN-SCAQRDLIQ----- 918
Db 803 IETTLTLANGNIGHIVILEGAQVNTTTTGTITIKVDNANANFSGTQTVTLIQGGARENG 862

QY 919 TLLIDSHDAGYARKMIDAT--SANE---ITKQLTATTTLLNNIASLEH----- 961
Db 863 TLGSPNFAVTGNSRNFVNSLIRAAQDYYVITRNAENVYVITNDIANSPFEGGAPGVQDQNYT 922

QY 962 -----KTSGLQTL-----SLSNAMILNLSRLNLSRRHTHDSIPAKRLQALKDORFASL 1010
Db 923 TPNVNTATNAAYNLLAKNSANSFVGAIVDTNSAITNVQLDLAKDIOAQLGNRLGAL 982

QY 1011 ESAAEVLVYQFAPKYEKPTNVWNAI-----GGTSLNN----- 1042
Db 983 R-----YLGTPETAEMAGPEAGASAAVAAGDEAIDNVAYGIWAKPFYVTAHQSKKGL 1036

QY 1043 -GSNASLYGTSAGVDAYLNGQVEAIVGGPGSYGYGSFNNRANSLNG-----ANNTNFGY 1097
Db 1037 AGYKAKTTGVVIGLDTLANDNL--MIG--AAIGITKTDIKHODYKKGDKTDVNGFSFSLY 1092

QY 1098 SRIFANOHEFDEACG-ALGSDQSSLSNFKSALLQDLNQSY-----HYLAYSAATPRSY 1149
Db 1093 G---AQQLVKNFFAOGSAIFSLNQVKNKSQRYFFDANGNWSKQIAAGHYDNMTFGNLTIV 1149

QY 1150 GYDFAFFRNALVKPSVGSYNHLGSLTFNPKSNSTNOVALKNGSS 1193
Db 1150 GYDYNAMQGVLV-TPMAGLSYLKSSDENYKETGTT-VANKOVNS 1191

RESULT 14
A41477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990

A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated, ne
A:Reference number: A41477; MUID:90354033
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: GB:M31227; NID:gl52465; PIDN:AAA26380.1; PID:gl52466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for residue
C:Keywords: surface antigen; tandem repeat
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 4.7%; Score 313; DB 2; Length 2249;
Best Local Similarity 19.5%; Pred. NO. 8.4e-08;
Matches 354; Conservative 193; Mismatches 589; Indels 676; Gaps 83;

Qy	14	LVSIALVGLSVTPOOSHAFPTTV-----IIPAIVGGIATGTAAGVTSGLLSUGLK	66
Dd	86	YGGDYTVTADAORIIKAINVACTPVGUNITQNTVVGSIIYKGNLLPVLNAGKSLFLN	145
Qy	91	-----NEPPKEYDLYRSL-----LSSKIDGWDWG----	116
Dd	146	GNNVAANHGFDPADNYTCLGNIALLGGANAALIIOAAPSKITLAGNIDGG--GIITV	202
Qy	117	--NAARHYWVKG--GOONKEVDMDAVGYTYLSG-----	147
Dd	203	KTDAA---INGTIGNTAL-ATPVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNAV	257
Qy	148	----LRNFTEGD--LDVN-MQKATRLRGFNGNSFTSYKDSA-----	182
Dd	258	LTGAIDNTTGDNVGVNLUNGALSQVTDIG--NTNSLATISVGAGTATLGGAVIKATT	315
Qy	183	---DRTRVDFENAKNI---SIDNFVEINRV-----GSGAGRKAASSVLT--L	222
Dd	316	KLTDAA SAVKFTNPVVVTGAIDNTGNANGIVTFGTGNSVTGNGVNTNALATPNVGC	375
Qy	223	QASEGTSDDKAIEISLYDGATL-----NLASSSVKLMGNVMWRLOYVG-----	266
Dd	376	QVOGGVV--KANTINTDNASAVTFNPNVVVTGAIDNTGNANGIVTFGTGNSVTGDI	433
Qy	267	-----AYLAPSYSTINTSKVTGEVFNHILTVGDKNAAOAGIANKK--TNIG	311
Dd	434	TNALATPNVVGAGTATLGGAVIKATTTKLTNAASV--LTUTNANAVLTGAIDNTTGD	491
Dd	312	TLDLWOSAGLNIIAPPEGGYKDPKNPTPSQS-----GAKNDKNESAKNDKPSS	360
Dd	492	VNL--NGALSQVGTGIG-----NTNSLATISVGAGTATLGGAVIKATTTKLTDA	542
Qy	361	QNNSNTQVINPPNSAOKTEVQPQTVIDG-----PFAGKDTVVNIRINTNADGT	411
Dd	543	VKTNPVVV-----TGAIDNTGNANGIVTFGTGNSVTGDIG--NTNSLATI	587
Qy	412	RVGGEKASL-----TTNAHLHGKGCVNLNSQAQRSLIVENLTGNITVDGPL	460
Dd	588	SVGAGTATLGGAVIKATTTKLTNAASV-----LTUTNANAVLTGAIDNTTGD	640
Qy	461	RVN---NQV--GGYALAGSSANFEFKATDTKNG-----TATFNNDISLGREV-	503
Dd	641	NLANGALSQVTDIGNTNSLATISVGAGTATLGGAVIKATTTKITNAVASAVKFTNP	700
Qy	504	-----NULKVDAMHTANFGKIDTNGCGFPNTDFSGVTDKVINKLITAST	546
Dd	701	AIDSTGNANNNGIVTFGTGNSVTGDI GNTNALATPNVVGAGTATLGGAVIKATT	760
Qy	547	NVAVKKFN--INELIVKTING---ISY---GEYTHFSESDIGSOSRI NTVRLETGR	597
Dd	761	VLTUTNANAVLTGAIDNTTGGDNVGVNLUNGALSQVTDIGINTNSLATISVGAGT	818
Qy	598	GGVKFK-----GGEKLVIDEFYYPSPNYYFOARNIK	627

Db	819	GGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGDNGVGLNL-----NGALSQVTTG	873
Qy	628	NVEITTKLAFPGOSPWGTSKL-----MFNNLTL-CONAVMDSYQSFNLTI	672
Db	874	DIGNTSLATISVGA--GPTATLGGAVIKATTTKLTNAASVLTLTNANAVLT-GAIDN-TT	929
Qy	673	QGFPIINOGTINY-----LVRGKVKATLSVGNAAAMFNNDIDSATGFKYKPLIKI	722
Db	930	GGD---NVGVLNLUNGALSQVTDGIGNTSLATISVGCAGTATLGGAVIKATT-----TKL	980
Qy	723	NSAODLIKNTHEVLLKAKLIIGXGNVSTG-----TNGLSNVNLEE--QF	763
Db	981	TDRASAVKFTNPVVVTCALDNTGNANGLVTFGTGNTVTGNGVNTNALTATVNGAGLLQV	1040
Qy	764	KERLALYNNNNRMDTC-VYRNTDDIKACGMAIGDOSVMYNNPDNYKYLIGKAKRNIGISK	822
Db	1041	QGGVVKANTINLTDNASAVTETNPVVVTTG-----AIDNT	1074
Qy	823	ANGSKISVYVLGNSTPENGNTNLPT-----NTTNSNAR-----	857
Db	1075	GNANNGIVTFTGNTSTVTGNGVNTNALTATVNVGAGLLQVGGVVKANTINLTDNASAVT	1134
Qy	858	-----SANNALA-----ONAPPAQPSAPNL-----V	879
Db	1135	NPVVVTCALDNTGNANGLVTFGTGNTVTGDTGNTNA-----LATVNVGAGITLQAGGSL	1189
Qy	880	AINQHDFG---TTESVFELANRSKOIDTLYANSQAQGRDLLQTLTLLIDSHDAGYARKMIDA	936
Db	1190	AANNIDFGARSTLEFNGPLDGGGKAIPYYFKCAIANGNAI-----	1230
Qy	937	TSANETIKQLNTATTTLNINIASLEHKTSGLQTL--SLSNAMILNLSRLVNLSSRRHTNIDS	994
Db	1231	--LNVNTKLLTASHLTIGTVAEINTGAGNLFTIDASGVDTILNAQNTINFRARDSVLV--	1286
Qy	995	FAKRLOALKDQKFASTESAEEVLYQFAPKYEKPTVWANAIGGTSLNNGSNASLYGTSG	1054
Db	1287	-----LSNLTGCVGNVNTLLAADLV---APCADEGTVFVNGCVNG--LVNGSIVA--GTARN	1335
Qy	1055	V-----DAYLNGQVEAIVGGFGSYGYS-SFNNRANLSNGS----	1088
Db	1336	IGDGGGNKFTLLIYNAVITTDVNLEGIONVLINKNADFTSSTAFNAGAIOINDATYTI	1395
Qy	1089	ANNTNFGYYS-----	1098
Db	1396	DANNGLNLTNPAGNIQFAHADAOLVLQNSGNDRTTTLGCANIDPDNDDEGIVILNSVTAGK	1455
Qy	1099	-----RIFANOHEFD---FEAQGALGSDOSSNFKSALLQDLNQ-----SVHYLAAYS	1142
Db	1456	KLPIAGGKFTGGAAHKLQTLTLFKGAGDCSTAGTFTNTIVLIDITGOELGATTANVVLFN	1515
Qy	1143	AATRASV-----GY-DFAPFRNALVKPSGVV--SYNHLGSTN-----FKSNSNQ	1185
Db	1516	DAVOLTOTGNIGGFLDFNAKNGMVTLLNNNVNAGAVQNTGTTNCTTLVILGASLNVRNG	1575
Qy	1186	VA-LKNGSSSOHL-----FNASNVAEARYYYGDT-S--FFYM	1218
Db	1576	IAMLKVGAGNVTIAKGGKYKIGEOGTGNTLTLPALHNLTGSIKNT---GGQALKLNFM	1632
Qy	1219	NAGVLOEFAHVGSNNAASLNTFKVNAARNPLTHARVMMGGELAKAEVFLNGLGVYLLHN	1278
Db	1633	NGGSVCVGTAAANSVDITTAGATSFASSVNAKGTATLGGTTSFA-NTFTNTCAVTLAK	1691
Qy	1279	LISNIGHFASNL	1290
Db	1692	--GSITSFAKNV	1701

RESULT 15

ABC-type transport protein ydbA.2 - Escherichia coli
C48399
C:Species: Escherichia coli
C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #
C:Accession: C48399; D64891; H64891

R: Moszer, I.; Glaser, P.; Danchin, A.
 Biochimie 73, 1361-1374, 1991
 A: Title: Multiple IS insertion sequences near the replication terminus in Escherichia coli
 A: Reference number: A48399; MUID: 92190338
 A: Accession: C48399
 A: Molecule type: DNA
 A: Residues: 484-2020 <MOS>
 A: Cross-references: GB:D85081; NID: g3041754
 A: Experimental source: strain K-12
 A: Note: sequence inconsistent with the nucleotide translation
 A: Note: sequence extracted from NCBI backbone (NCBIN:88089, NCBI:88090)
 R: Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A: Title: The complete genome sequence of Escherichia coli K-12.
 A: Reference number: A64720; MUID: 97426617
 A: Accession: D64891
 A: Status: nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-839, 'LDLPLYFQTSVIT' <BLA1>
 A: Cross-references: GB:AF000237; GB:U00096; NID: g1787665; PID: A64720.1; PID: g1787668;
 A: Experimental source: strain K-12, substrain M61655
 A: Accession: H64891
 A: Status: nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 'M', 915-2020 <BLA2>
 A: Cross-references: GB:AF000237; GB:U00096; NID: g1787665; PID: A64720.1; PID: g1787672;
 A: Experimental source: strain K-12, substrain M61655
 C: Genetics:
 A: Gene: ydbA_2
 A: Start codon: GTG

Query Match 4.7%; Score 312.5; DB 2; Length 2020;
 Best Local Similarity 21.2%; Pred. No. 7.6e-08;
 Matches 320; Conservative 192; Mismatches 574; Indels 423; Gaps 79;

Qy 50 TGTAVGTGSLGSLGK-----QAEENKTPDKPKVRIQAGKGFNEFPNKEDLYRS 103
 Db 165 TENADGTISLQDSNGRKATINLWQIDEANTV-----A 197

Qy 104 LLSKIDGGDWG--NAARHYVWG-----GOONKLEVDKMAVGTYTSLGLRN--FTGGD 155
 Db 198 LEGVSADGATKQYVNHGELVITGDNATVNNNGKTTVDGKSTGT--EINGNNGKVIQDGD 256

Qy 156 LDVNMKATLRGQFNGSFTSYKDSADR---TTRVDFNAKNISIDNFEVNNRVG---- 208
 Db 257 LDVS-----GGHGIDITGDSATVNDKGTMTVDPESMCIQIDGDKAIVNNEGESTI 308

Qy 209 --SGAGRKAS-----STVLTLQASEG--ITSKNAEIS--LYDGATLNL 246
 Db 309 TNGGTGTQINGDDATANNNGKTTVDGKSTGTGEINGNNGKVIQDGLDVSOGGHGIDITG 368

Qy 247 ASSSVKLMGNVWMLQVAYLAPSYTINT-----SKVTGE-----VNFNHL 290
 Db 369 DSATVDNKGTTMTVDPESIGVQDQAVVNEGESAITNGGTGTQINGDDATANNNGKT 428

Qy 291 TVGDKNAQAAGLIAN--KKTNIGTLDLQASAGLNIAPPEGYKDKPN---NTPSQSA 344
 Db 429 TVDGKDSGTETIAGNNGKVIQDGLDV--SGGGHGIDITGDSATVNDKGTMTVDPESIGI 487

Qy 345 KNDKNESAKNDKQESSQNSNTQV--IN----PPNSAQKTEVQPTQVIDGPFAGGRDVTVN 399
 Db 488 QIDGQAIIVNNEGESTITNGGTGTQINGDDATANNNGKT-----TVDGKDSGTGTKIAGN 541

Qy 400 INRINTNADGTRVGGFKASLTNAHLHGKGGVNLNQASGRSLIVENTLGNITV---- 456
 Db 542 IGVIV--NLDGSLTV-----TGGAH---GVENIGDNGTVNNKGDIVVSDTGSIGVLIN 588

Qy 457 -----DGPLRVNNQVGGY-----ALAGSSANFEKAGTDTKNGTATFNNDLSL 499
 Db 589 GEGATVNTGDNVNSNEATGFSITTNNSKVSLSAGSMQVGFSTGVDL--NGN---NNSVTL 644

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:28:49 ; Search time 18.18 Seconds

(without alignments)

2760.206 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	6671	100.0	1296	1	VACL_HELPY	Q48247	helicobacte
2	6303.5	94.5	1287	1	VACL_HELPY	Q48245	helicobacte
3	6222.5	93.2	1291	1	VAC4_HELPY	Q48258	helicobacte
4	6220	93.2	1290	1	VAC4_HELPY	P55981	helicobacte
5	6108	91.6	1288	1	VAC3_HELPY	Q92kw5	helicobacte
6	5237	78.5	1310	1	VAC3_HELPY	Q48253	helicobacte
7	335.5	5.0	1654	1	OMP8_RICRI	O53047	r outer mem
8	335	5.0	1655	1	OMP8_RICCN	Q9Kka3	r outer mem
9	330.5	5.0	1656	1	OMP8_RICJA	O06653	r outer mem
10	326	4.9	2003	1	YDBA_ECOLI	P33666	eschericchia
11	313	4.7	2249	1	OMPA_RICRI	P15921	rickettsia
12	312.5	4.7	1300	1	12OK_RICRI	P14914	rickettsia
13	310.5	4.7	2021	1	OMPA_RICCN	O52657	rickettsia
14	310	4.6	1286	1	ATDA_ECOLI	Q03155	eschericchia
15	306	4.6	1643	1	OMP8_RICRP	Q53020	r outer mem
16	291.5	4.4	1953	1	BTGA_SALTY	P25927	salmonella
17	291	4.4	1645	1	OMP8_RICTY	P66989	r outer mem
18	283	4.2	1541	1	IGA1_HAEIN	P42782	haemophilus
19	280	4.2	1325	1	YDEK_ECOLI	P32051	eschericchia
20	276	4.1	1545	1	IGA3_HAEIN	P43385	haemophilus
21	265.5	4.0	1702	1	IGA2_HAEIN	P45384	haemophilus
22	262.5	3.9	1569	1	YFJA_ECOLI	P52143	eschericchia
23	261.5	3.9	1694	1	IGA0_HAEIN	P44969	haemophilus
24	253.5	3.8	1460	1	PMPC_CHLMU	Q9p1y1	chlamydia m
25	246	3.7	917	1	HXA3_HAEIN	P45355	haemophilus
26	238	3.6	1394	1	HAP_HAEIN	P45387	haemophilus
27	235.5	3.5	1276	1	PMPE_CHLPN	Q92899	chlamydia p
28	219.5	3.3	2710	1	TOXA_CLODI	P16154	clostridium
29	217.5	3.3	1341	1	VG37_BPT2	P07067	bacterioph
30	217.5	3.3	1849	1	IGA4_HAEIN	P45386	haemophilus
31	213.5	3.2	1531	1	PMPC_CHLTR	O84818	chlamydia t
32	213.5	3.2	1577	1	HLVA_PROMI	P16466	proteus mir
33	213	3.2	918	1	YMJB_CAEEL	P34487	caenorhabdi

P39180 escherichia
P47551 mycoplasma
O84418 chlamydia t
P47502 mycoplasma
O92812 chlamydia p
P09790 neisseria g
Q07833 bacillus su
P45354 haemophilus
O84419 haemophilus
P15320 serratia ma
Q9jyv5 neisseria m
P49331 streptococ

ALIGNMENTS

```
RESULT 1
VACL_HELPY
ID VACL_HELPY STANDARD; PRT; 1296 AA.
AC Q48247; Q53434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=94222514; PubMed=8168917;
RA Phadnis S.H., Ilver D.J., Jansson L., Normark S., Westblom T.U.;
RT "Pathological significance and molecular characterization of the
RL vacuolating toxin gene of Helicobacter pylori.";
RL Infect. Immun. 62:1557-1565(1994).
RN [2]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=94216933; PubMed=8163943;
RA Telford J.L., Ghiara P., Dell'Orco M., Comanducci M., Burroni D.,
RA Bugnoli M., Tecce M.F., Censini S., Covacci A., Xiang Z., Papini E.,
RA Montecucco C., Parente L., Rappuoli R.;
RT "Gene structure of the Helicobacter pylori cytotoxin and evidence of
its key role in gastric disease.";
RL J. Exp. Med. 179:1653-1658(1994).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC -!- ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U07145; AAA18867.1; -
CC EMBL; S72494; AAB30582.1; -
CC InterPro; IPR003842; VACA.
CC Pfam; PF02691; VACA; 1.
CC Cytotoxin; Toxin; Signal.
CC SIGNAL 1 33 POTENTIAL.
CC CHAIN 34 ? VACUOLATING CYTOTOXIN.
CC PROPEP ? 1296 POTENTIAL.
SQ SEQUENCE 1296 AA; 139760 MW; 0D1F3F71AB411447 CRC64;
```

Query Match 100.0%; Score 6671; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 0; Indels 0; Gap

Qy 1 MEIOOTHRKINRPLVSLALVGVLSITPQQSHAAFFTTTIIPIAVGGTATGTVTSGL 60
Db 1 MEIOOTHRKINRPLVSLALVGVLSITPQQSHAAFFTTTIIPIAVGGTATGTVTSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120
Qy 121 HYWKGGQONKLEVDMDKAVGTYTSLGLRNFPGDGLDVMQKATLRGLQFNGNSFTSYKD 180
Db 121 HYWKGGQONKLEVDMDKAVGTYTSLGLRNFPGDGLDVMQKATLRGLQFNGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNISIDNFEYINRNVGSGAGRKASSTVLTQLQASEGITSKNAEISLYD 240
Db 181 SADRTTRVDFNAKNISIDNFEYINRNVGSGAGRKASSTVLTQLQASEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Qy 301 GIIANKTINICTLWOSAGLNIITAPPEGGYKDKPNNTPSOSGAKNDKNSAKNDKQESS 360
Db 301 GIIANKTINICTLWOSAGLNIITAPPEGGYKDKPNNTPSOSGAKNDKNSAKNDKQESS 360
Qy 361 QNNSNTQVNPNSAQKTEVQTVQIDGPPFAGGKDTVVVNIINRINTNADGTTIRVGGFKASL 420
Db 361 QNNSNTQVNPNSAQKTEVQTVQIDGPPFAGGKDTVVVNIINRINTNADGTTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGVNLSNOAGSRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGVNLSNOAGSRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNLTDFSGVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNLTDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTETRSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIGSQSRINTVRLTETRSLFSGGV 600
Qy 601 KFKGGEKLVIDFYYSPHNVYDARINKVETLNKLAGCPGSPWCTSKLMFNNTLQONA 660
Db 601 KFKGGEKLVIDFYYSPHNVYDARINKVETLNKLAGCPGSPWCTSKLMFNNTLQONA 660
Qy 661 VMDYSQFSLNLTQGDFFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLNLTQGDFFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Qy 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
Qy 781 VRNTDDIKACGMATGDSQSVWNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACGMATGDSQSVWNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Qy 841 NGGNTNLPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGTIESVFELANRSK 900
Db 841 NGGNTNLPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGTIESVFELANRSK 900
Qy 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Db 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Qy 961 HKTSGLOTLTSLNAMILNSRLVNSRRHTNHIDSFARKLQALKOOKPASLESAAEVLQOF 1020
Db 961 HKTSGLOTLTSLNAMILNSRLVNSRRHTNHIDSFARKLQALKOOKPASLESAAEVLQOF 1020
Qy 1021 APKYEKPTNVWANAIGTGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSYSSPNN 1080
Db 1021 APKYEKPTNVWANAIGTGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSYSSPNN 1080

Qy 1081 RANSLNSGANTNTEGVYSRIFANQHEPDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Db 1081 RANSLNSGANTNTEGVYSRIFANQHEPDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Qy 1141 YSAATRASGYDEFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1141 YSAATRASGYDEFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Qy 1201 SANVEARYYYGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Qy 1261 LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF 1296

RESULT 2

VAC2_HELPY
ID VAC2_HELPY STANDARD: PRT; 1287 AA.
AC Q48245;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN (1)
SEQUENCE FROM N.A.
RC STRAIN=ATCC 49503 / 60190;
RX MEDLINE=94193753; PubMed=8144644;
RA Cover T.L., Tummuru M.K., Cao P., Thompson S.A., Blaser M.J.;
RT "Divergence of genetic sequences for the vacuolating cytotoxin among
Helicobacter pylori strains.";
RL J. Biol. Chem. 269:10566-10573(1994).
CC -1- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U05676; AAA17657.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 33
FT CHAIN 34 ?
FT PROPEP 1287
SQ SEQUENCE 1287 AA; 139041 MW; 0007370062FCB71F CRC64;

Query Match 94.5%; Score 6303.5; DB 1; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;

Qy 1 MEIOOTHRKINRPLVSLALVGVLSITPQQSHAAFFTTTIIPIAVGGTATGTVTSGL 60
Db 1 MEIOOTHRKINRPLVSLALVGVLSITPQQSHAAFFTTTIIPIAVGGTATGTVTSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLRSLSSKIDGGWDMGNAAR 120
Qy 121 HYWKGGQONKLEVDMDKAVGTYTSLGLRNFPGDGLDVMQKATLRGLQFNGNSFTSYKD 180


```
Db 121 HWIRGGQWNLKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSDKNAEISLYD 240
Db 181 SADRTTRVDENAKNIIIDNFIENNRVSGAGRKASSVTLTQASSEGITSSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Qy 301 GIIANKKTNIGTLDLWQASGLNIIAPPGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS 360
Db 301 GIIASNKTHIGTLDLWQASGLNIIAPPGGYKDKPNNTPSQSG-----AKNDKQESS 352
Qy 361 QNNSNTQVINPNNSAKTQEVQTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVINPNNSAKTQEVQTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGGVNLNOASGRSLIVENLTGNITVDGPLRVNQVGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGGVNLNOASGRSLIVENLTGNITVDGPLRVNQVGYALAGSSANFEF 472
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAAHTANFKGIDTNGGPFNTLDFSGVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDAAHTANFKGIDTNGGPFNTLDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNELIVKNTGTSVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 600
Db 541 LITASTNVAVKFNELIVKNTGTSVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 600
Qy 601 KFKGGEKLIVDEFYYSWPNYEDARNIKVEITNKLAFGPGQSPWCTSKLMFNLLTGONA 660
Db 601 KFKGGEKLIVDEFYYSWPNYEDARNIKVEITNKLAFGPGQSPWCTSKLMFNLLTGONA 652
Qy 661 VMDYSQFNLTIQGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 661 VMDYSQFNLTIQGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 712
Qy 721 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 772
Qy 781 VRNTDDIKACGMAIGDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACGMAIGDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 832
Qy 841 NGGNTNLTPTNTTSNARSANALQANAPPAQPSATPNLVAINQHDGFTIESVFELANRSK 900
Db 841 NGGNTNLTPTNTTSNARSANALQANAPPAQPSATPNLVAINQHDGFTIESVFELANRSK 891
Qy 901 DIDTLYANSAGQGRDLQTLTLDSDHAGYARKMIDATSAEITTKOLNTATTTLNNAISLE 960
Db 901 DIDTLYANSAGQGRDLQTLTLDSDHAGYARKMIDATSAEITTKOLNTATTTLNNAISLE 951
Qy 961 HKTSGLOTLISLNSAMILNSRLNLSRRHTNHDSPAKRLQALQKQKQFASLESAAEVLQF 1020
Db 961 HKTSGLOTLISLNSAMILNSRLNLSRRHTNHDSPAKRLQALQKQKQFASLESAAEVLQF 1011
Qy 1021 APKYEKPTNVNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFSGYSFNN 1080
Db 1021 APKYEKPTNVNANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFSGYSFNN 1071
Qy 1081 RANLSNGANTNFCVYSKRIFANOHEPDEAQQALGDSQSSLNFKSALLQDLNQSYYHLA 1140
Db 1081 RANLSNGANTNFCVYSKRIFANOHEPDEAQQALGDSQSSLNFKSALLQDLNQSYYHLA 1131
Qy 1141 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1141 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1191
Qy 1201 SANVEARYYYGDTSYFYMNAGVQLQFAHVHGSNNAAASLNTFKVNAARNPLNTHARVMMGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVQLQFAHVHGSNNAAASLNTFKVNAARNPLNTHARVMMGE 1251
```

```
Db 1192 SANVEARYYYGDTSYFYMNAGVQLQFAHFANFGSSNAVSLNTFKVNATRNPLNTHARVMMGE 1251
Qy 1261 LKLAKEVFLNLGVVYVHLNLIISNIGHFASNLGMRYSF 1296
Db 1252 LKLAKEVFLNLGVVYVHLNLIISNIGHFASNLGMRYSF 1287

RESULT 3
ID VAC4_HELPY STANDARD; PRT: 1291 AA.
AC Q48258;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=185-44;
RX MEDLINE=94333650; PubMed=8057855;
RA Haas R., Schmitt W.;
RT "Genetic analysis of the Helicobacter pylori vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein."
RL Mol. Microbiol. 12:307-319(1994).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)
CC -----
DR EMBL; Z26883; CAA81528.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 33
FT CHAIN 34 ?
FT PROPEP ? 1291
FT POTENTIAL.
SQ SEQUENCE 1291 AA; 139635 MW; ECA56A61CAE36669 CRC64;

Query Match 93.3%; Score 6222.5; DB 1; Length 1291;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1202; Conservative 51; Mismatches 38; Indels 5; Gaps 1;

Qy 1 MEIQOTHRKINRPLVSLALVGVLSITPOQSHAAFTVTIIPAIYGGIATGATVGTSGL 60
Db 1 MEIQOTHRKINRPLVSLALVGVLSITPOQSHAAFTVTIIPAIYGGIATGATVGTSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEDLYKSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEDLYKSLSSKIDGGWDMGNAAR 120
Qy 121 HWYKGGQONKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HWYKGGQONKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSDKNAEISLYD 240
Db 181 SADRTTRVDENAKNIIIDNFIENNRVSGAGRKASSVTLTQASSEGITSSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
```

```
Db 241 GATLNLSASVVKLMGNVWVGRLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDHNAQA 300
Qy 301 GIANKTKTNGTLDLWQAGLNIIPAPGGYKDKPNTTPSQGAKNDKESAKNDKQESS 360
Dl 301 GIASNKTHTGTLWQAGLNIIPAPGGYKDKPNDKPS-----NTQNNAKNDKQESS 355
Qy 361 QNNSNTQVIPPNSAQTEQVQTOVDGPPAGGKDTVVNINRINTNADGTRVGGFKASL 420
Db 356 QNNSNTQVIPPNSAQTEQVQTOVDGPPAGGKNTVVNINRINTNADGTRVGGFKASL 415
Qy 421 TTNAHLHIGKGVNLSQASGRSLIVENLTGNTVDGPLRVNNOVGAYLAGSNAFEE 480
Db 416 TTNAHLHIGKGVNLSQASGRSLIVENLTGNTVDGPLRVNNOVGAYLAGSNAFEE 475
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATANKFGIDTGNNGENTLDFSGVTDKVNIN 540
Db 476 KAGTDTKNGTATFNNDISLGRFVNLKVDATANKFGIDTGNNGENTLDFSGVTDKVNIN 535
Qy 541 LITASTNVAVKFNINELIVKTNISVGEYTHFSEIDIGSQRINTVYLETGTRSLFSGV 600
Db 536 LITASTNVAVKFNINELIVKTNISVGEYTHFSEIDIGSQRINTVYLETGTRSIYSGV 595
Qy 601 KFKGGEKLVIDEFYSPNVEDARNIKNVEITNKLAFPOGSPHGTSKLMFNNTLGQNA 660
Db 596 KFKGGEKLVIDEFYSPNVEDARNIKNVEITNKLAFPOGSPHGTAKLMFNNTLGQNA 655
Qy 661 VMDYSQFSLNTIQGDFINNOGTINYLVRGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 656 VMDYSQFSLNTIQGDFINNOGTINYLVRGKVATLSVGNAAAMFFSNVDSATGYKPLI 715
Qy 721 KINSADLIKNTHEVLKAKIIGYGVNSTGTNGISNVLBEEQKRLALYNNNRMDTCV 780
Db 716 KINSADLIKNTHEVLKAKIIGYGVNSTGTNGISNVLBEEQKRLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMALGDSQWNNPDNYKYLIGKAKWNGISKTANGSKISVYILGNSTPTE 840
Db 776 VRNTDDIKACGMALGDSQWNNPDNYKYLIGKAKWNGISKTANGSKISVYILGNSTPTE 835
Qy 841 NGGNTNLPNTTNSARNSANALQAAPFAQPSATPNLVAINHDFGTIESVFELANRSK 900
Db 836 NGGNTNLPNTTNSARNSANALQAAPFAQPSATPNLVAINHDFGTIESVFELANRSK 895
Qy 901 DIDTLANSAGQGRDILLQTLIDSHDAGYARKMIDATSAEITKOLMTATTLNNAISLE 960
Db 896 DIDTLANSAGQGRDILLQTLIDSHDAGYARKMIDATSAEITKOLMTATTLNNAISLE 955
Qy 961 HKTSGLOTLSLSNAMILNSRLVNLRRHTNHDISFAKRLQALKDQKFALESAAEVLYQF 1020
Db 956 HKTSGLOTLSLSNAMILNSRLVNLRRHTNHDISFAKRLQALKDQKFALESAAEVLYQF 1015
Qy 1021 APYKPTNYWANAIGTSLNNGSNASLYGTSAGVDAYLNGQVBAIVGGFSGYSFNN 1080
Db 1016 APYKPTNYWANAIGTSLNNGSNASLYGTSAGVDAYLNGQVBAIVGGFSGYSFNN 1075
Qy 1081 RANSLNSGANTNFGVYSRIFANQHEDFEAQAGLSDQSLSNFKSALLQDLNQSYYHLA 1140
Db 1076 RANSLNSGANTNFGVYSRIFANQHEDFEAQAGLSDQSLSNFKSALLQDLNQSYYHLA 1135
Qy 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKNSNTNOVALKNGSSSOHLFNA 1200
Db 1136 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKNSNTNOVALKNGSSSOHLFNA 1195
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHGVSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFARHGVSNNAASLNTFKVNAARNPLNTHARVMMGGE 1255
Qy 1261 LKLAKEVFLNLGVYVYLNHLNLSNTHGFASNLGMRYSF 1296
Db 1256 LKLAKEVFLNLGVYVYLNHLNLSNTHGFASNLGMRYSF 1291
```

RESULT 4
VACA_HELPY

```
ID VACA_HELPY STANDARD; PRT; 1290 AA.
AC P55981;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA OR HP0887.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000598; AAD07935.1; -;
DR TIGR: HP0887; -;
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal; Complete proteome.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 ? VACUOLATING CYTOTOXIN.
FT PROPEP 1290 AA; 139312 MW; F48B23513447A1AC CRC64;
FT SEQUENCE 1290 AA; 139312 MW; F48B23513447A1AC CRC64;
Query Match 93.2%; Score 6220; DB 1; Length 1290;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1207; Conservative 38; Mismatches 45; Indels 6; Gaps 2;
Qy 1 MEIQOTHRKINRPLVSLALVGLVSIPTQSSHAAFFTTVIIPAVGGIATGAVGTVSGL 60
Db 1 MEIQOTHRKINRPLVSLALVGLVSIPTQSSHAAFFTTVIIPAVGGIATGAVGTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNFPNKEVDLYRSLSSKIDGWDGNGAAR 120
Db 61 LGWGLKQAEANKTPDKPKVRIQAGKGFNFPNKEVDLYRSLSSKIDGWDGNGAAR 120
Qy 121 HYWYKGGQKQKLEVDKMDKAVGTYYTLGLRNFTGGDLVDNMQKATLRLGQFNGNSFTSYKD 180
Db 121 HYWYKGGQKQKLEVDKMDKAVGTYYTLGLRNFTGGDLVDNMQKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNISIDNFEINNRVCGSAGRKASSTVLTQLASEGITSKNAEISLYD 240
Db 181 SADRTTRVDFNAKNILIDNFEINNRVCGSAGRKASSTVLTQLASEGITSKNAEISLYD 240
Qy 241 GATLNLSASVVKLMGNVWVGRLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDHNAQA 300
Db 241 GATLNLSASVVKLMGNVWVGRLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDHNAQA 300
```

QY 301 GIANKKTIGTLDLWQSLNIIAPPEGGYKDKPNTPSQGAKNDKESAKNDKQESS 360
DB 301 GIANSKTHIGTLDLWQSLNIIAPPEGGYKDKPKDPS-----NTTQNNANNQNSA 355
QY 361 QNNSNTQVNPNSAQKTEVQPTVIDGPFAGKDTVVNNINRINTNADGTTIRVGGFKASL 420
DB 356 QNNSNTQVNPNSAQKTEIQPTVIDGPFAGKDTVVNNIDRINTNADGTTIRVGGFKASL 415
QY 421 TTNAAHLHIGKGVNLNOAGSRSLIVENLTGNITVDGPLRVNNOVGGYALAGSSANPEF 480
DB 416 TTNAAHLHIGKGVNLNOAGSRLLVLENLTGNITVDGPLRVNNOVGGYALAGSSANPEF 475
QY 481 KAGTDTKNGTATFNNDISLGRFVNKLVDKDAHTANFKGIDTNGGFGNTLDFSGVTDKVNINK 540
DB 476 KAGTDTKNGTATFNNDISLGRFVNKLVDKDAHTANFKGIDTNGGFGNTLDFSGVTDKVNINK 535
QY 541 LITASTNVAKVFNNELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLFSGGV 600
DB 536 LITASTNVAKVFNNELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLFSGGV 595
601 KFKGSEKLVDFEYSPWNYFDARNIKNVEITNKLAFGPQGGPWGTSKLMFNNTLQONA 660
596 KFKGSEKLVDFEYSPWNYFDARNIKNVEITRKFASTPENPWGTSKLMFNNTLQONA 655
QY 661 VMDYSQFSLTIQGDFFINNQGINTVLVRGKGVATLSVGNAAAMFNNDIDSATGYKPLI 720
DB 656 VMDYSQFSLTIQGDFFINNQGINTVLVRGQVATLVGNAAAMFNNDIDSATGYQPLM 715
QY 721 KINSQDLKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEDEQFKERLALYNNNRMDTCV 780
DB 716 KINSQDLKNTHEVLLKAKIIGYGNVSTGTNSISNVNLEIEQFKERLALYNNNRMDICV 775
QY 781 VNTDDIKACMAIGDQSVNPNPNKYLLIGKAWKNIGTSKTANGSKISVYVLGNSTPTE 840
DB 776 VNTDDIKACGTAIGQSVNPNPNKYLLIGKAWKNIGTSKTANGSKISVYVLGNSTPTE 835
QY 841 NGGNTTNLTPTNTSARSANALQAQPAQPSATPNLVAIQHDFGTIESVFELANRSK 900
DB 836 KGGNTTNLTPTNTSARSANALQAQPAQPSATPNLVAIQHDFGTIESVFELANRSK 895
QY 901 DITDLYANSAGORLLQTLTLDSDHAGYARKMIDATSAEITKQLNTATTLNNTASLE 960
DB 896 DITDLYANSAGORLLQTLTLDSDHAGYARQMDTSTGEITKQLNAAATTLNNTASLE 955
QY 961 HKTSGLOTLSLNNMLNRLNRRHTNHIDSPAKRLQALKDQKFALESAAEVLQF 1020
DB 956 HKTSSLOTLSLNNMLNRLNRRHTNHIDSPAKRLQALKDQKFALESAAEVLQF 1015
QY 1021 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSSPNN 1080
1016 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSSPNN 1075
QY 1081 RANSLNGANNNTFGYVSRIFANQHEFDEAOCALGSDQSSLNFKSALLQDLNQSYHYLA 1140
DB 1076 QANSLNGANNNTFGYVSRIFANQHEFDEAOCALGSDQSSLNFKSALLQDLNQSYHYLA 1135
QY 1141 YSAATRASGYDFAFRNALVLKPSGVSYNHLGTSNFKSNSTNOVALKNGSSQHLFNA 1200
DB 1136 YSAATRASGYDFAFRNALVLKPSGVSYNHLGTSNFKSNSTNOVALKNGSSQHLFNA 1194
QY 1201 SANVEARYYYGDTSYFYMAGVLQEFARHYSNNAASLNTFKYNAARNPLNTHARVMGGE 1260
DB 1195 SANVEARYYYGDTSYFYMAGVLQEFANFGSSNAVSLNTFKYNAARNPLSTHARVMGGE 1254
QY 1261 LKLAKEVFLNLGVYVHLNLSINIGHFASNLGMYSF 1296
DB 1255 LKLAKEVFLNLGVYVHLNLSINIGHFASNLGMYSF 1290

RESULT 5
VACA_HELPJ
ID VACA_HELPJ STANDARD; PRT; 1288 AA.

Q92KW5;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA OR JHP0819.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001511; AAD06400.1; -;
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal; Complete proteome.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 ? VACUOLATING CYTOTOXIN.
FT PROPEP ? 1288 POTENTIAL.
SQ SEQUENCE 1288 AA; 139131 MW; 2448159DFC5F32B9 CRC64;

Query Match 91.6%; Score 6108; DB 1; Length 1288;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1181; Conservative 53; Mismatches 54; Indels 8; Gaps 2;

QY 1 METQOTHRINRPLVSLALVGVSTPQOSHAAFTTIIIPAVGGIATGTAVGTVSG 60
DB 1 METQOTHRINRPLVSLVLAGALISAIPQESHAAFTTIIIPAVGGIATGTAVGTVSG 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDGNAAR 120
DB 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDGNAAR 120
QY 121 HWYKGGQGNKLEVDKMDKAVGTYTSLGLRNFSGDLDVNNQKATRLGQFNNGSFTSYKD 180
DB 121 HWYKGGQGNKLEVDKMDKAVGTYKLSGLRNFSGDLDVNNQKATRLGQFNNGSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFEINNRVSGAGRKASSTVLTLOASEGITSSKNAEISLYD 240
DB 181 SADRTTRVDFNAKNTSIDNFEINNRVSGAGRKASSTVLTLOASEGITSSKNAEISLYD 240
QY 241 GATNLASSSVKLMGVNWMGRLOQYVAYLAPSYSTINTSKYTCGEVFNHLYTGDKNAAQA 300
DB 241 GATNLASSSVKLMGVNWMGRLOQYVAYLAPSYSTINTSKYTCGEVFNHLYTGDKNAAQA 300
QY 301 GIANKKTIGTLDLWQSLNIIAPPEGGYKDKPNTPSQGAKNDKESAKNDKQESS 360
DB 301 GIANSKTHIGTLDLWQSLNIIAPPEGGYKDKPNSTTSQSGTKNDKKEISQN----- 354
QY 361 QNNSNTQVNPNSAQKTEVQPTVIDGPFAGKDTVVNNINRINTNADGTTIRVGGFKASL 420
DB 361 QNNSNTQVNPNSAQKTEVQPTVIDGPFAGKDTVVNNINRINTNADGTTIRVGGFKASL 420

Db 355 - NNSNTEVINPPNNTQKTETPTQVIDGPFAGGKDTVVNIFHLNKTADGTTIKVGGFKASL 413
Qy 421 TTNAHLHIGKGGVNLNOAGSRSLVNLGNITVDGPLRVNNOVGAYALAGSANPEF 480
Db 414 TTNAHLNIGKGGVNLNOAGSRLLVNLGNITVDGPLRVNNOVGAYALAGSANPEF 473
Qy 481 KAGDTDKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNCGFNTLDFSGVTDKVNINK 540
Db 474 KAGVTDKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNCGFNTLDFSGVTDKVNINK 533
Qy 541 LITASTNAVAVNFNINELIVTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 600
Db 534 LITASTNAVAVNFNINELIVTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 593
Qy 601 KFKGKELVIDEFYSPNYPNADNRKINVEITNKLAFQPGQSPWCTSKLMFNNTLQONA 660
Db 594 KFKGKELVIDEFYSPNYPNADNRKINVEITNKLAFQPGQSPWCTSKLMFNNTLQONA 653
Qy 661 VMDYSQFSLTIQDGFINNQTINYLVRGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 654 VMDYSQFSLTIQDGFINNQTINYLVRGKVATLSVGNAAAMFNNDIDSATGYKPLI 713
Qy 721 KINSADQLIKTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 714 KINSADQLIKTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 773
Qy 781 VRNTDDIKACGMAIGDQSMVNNPNYKYLIKAWNIGISKTAGSKISVYVLGNSTPTE 840
Db 774 VRNTDDIKACGMAIGDQSMVNNPNYKYLIKAWNIGISKTAGSKISVYVLGNSTPTE 833
Qy 841 NGGNTNLPNTTNSARSNALNAQNPAPPAQSPATPNLVAINQHDFTGIESVFELANRSK 900
Db 834 NGGNTNLPNTTNSARSNALNAQNPAPPAQSPATPNLVAINQHDFTGIESVFELANRSK 892
Qy 901 DIDTLYANGAAGRDLLQTLIDSHDAGYARKMIDATSAEITKOLNTATTTLNNAISLE 960
Db 893 DIDTLYTHSGAAGRDLLQTLIDSHDAGYARKMIDNTSTGEITKOLNAATDALNNAISLE 952
Qy 961 HKTSGLOTLSLSNAMILNLSRLVNLRRHTNIDSFARLQALQKQKFALESAAEVLQF 1020
Db 953 HKQSGLOTLSLSNAMILNLSRLVNLRRHTNIDSFARLQALQKQKFALESAAEVLQF 1012
Qy 1021 APKYEKPTNVWANAIGTSLNNGSNASLYGTSGAGDAYLNGQVATVGGFSGSYGSFNN 1080
Db 1013 APKYEKPTNVWANAIGTSLNNGSNASLYGTSGAGDAYLNGQVATVGGFSGSYGSFNN 1072
Qy 1081 RANSLGANNNTNFGVYSRIFANQHEFDEAQAAGLQDQSSLNFKSALLQDLNOSYHYLA 1140
Db 1073 QANSLGANNNTNFGVYSRIFANQHEFDEAQAAGLQDQSSLNFKSALLQDLNOSYHYLA 1132
Qy 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1133 YSARASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1192
Qy 1201 SANVEARYYYGDTSYFYFNAAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
Db 1193 NANVEARYYYGDTSYFYFNAAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1252
Qy 1261 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMWYSF 1296
Db 1253 LQLAKEVFLNLGVVYVHLNLSNIGHFASNLGMWYSF 1288
RESULT 6
VAC3_HELPY
ID VAC3_HELPY STANDARD; PRT; 1310 AA.
AC Q48253;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TX30A;
RX MEDLINE=95355366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vac types with cytotoxin production and
RT peptic ulceration.";
RL J. Biol. Chem. 270:17771-17777(1995).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U29401; AAA86834.1; .
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca: 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 ? VACUOLATING CYTOTOXIN.
FT PROPEP ? 1310 POTENTIAL.
SQ SEQUENCE 1310 AA; 141988 MW; 1BC21FE3D435F981 CRC64;

Query Match 78.58; Score 5237; DB 1; Length 1310;
Best Local Similarity 77.98; Pred. No. 6.2e-270;
Matches 1045; Conservative 88; Mismatches 131; Indels 78; Gaps 13;

Qy 1 MEIOOTHRKINRPLVSLALVGLVSI-----TP-----QQSHAAFTFTVIIIPAIYGGIATG 51
Db 1 MEIOOTHRKINRPLVSLALVGLVGLMTELGANTPNPHSESRATFTVIIIPAIYGGIATG 60

Qy 52 TAVGTVSGLLSWGLQKQAEANKTPDKPVWRIQAGKGFNEFPNKEYDLYSLSSKIDG 111
Db 61 AAVGTVSGLLSWGLQKQAEANKTPDKPVWRIQAGKGFNEFPNKEYDLYSLSSKIDG 120

Qy 112 GWDHGNARHYVWVGQONKLEVDKDAVGTYTLGSLRNFTHGGDLVNMOKATLRLGQFN 171
Db 121 GWDHGNARHYVWVGQONKLEVDKDAVGTYTLGSLRNFTHGGDLVNMOKATLRLGQFN 180

Qy 172 GNSFTSYKDSADRTVRVDFNAKNSIDNFVEINNRVSGAGKASSTVLTQLQASEGITS 231
Db 181 GNSFTSYKDSADRTVRVDFNAKNSIDNFVEINNRVSGAGKASSTVLTQLQASEGITS 240

Qy 232 KNAEISLYDGTATLNLAS-----SVKLMGNVWGRLOYVGYLAPSYNTINTSKVTGEVNFN 288
Db 241 ENAEISLYDGTATLNLSSNQSDVLYGKVMGRLOYVGYLAPSYNTINTSKVTGEVNFN 300

Qy 289 HLTVDKNAOAGIANKKTNIGTLQWSAGLNTIAPPEGYKDKPNNTPSQSGAKNDK 348
Db 301 HLTVDKNAOAGIANKKTNIGTLQWSAGLNTIAPPEGYKDKPNNTPSQSGAKNDK 351

Qy 349 NESAKNDKQESSNNSTQVFNPPNSAOKTEVQPTQVIDGPFAGGKDTVVNINRINTNAD 408
Db 352 -----QN-----NPKNDAGKTEIQPTQVIDGPFAGGKDTVVNIFHLNTRAD 392

Qy 409 GTIRVGGFKASLTNAHLHIGKGGVNLNOAGSRSLVNLGNITVDGPLRVNNOVG 468
Db 393 GTIRVGGFKASLTNAHLHIGKGGVNLNOAGSRSLVNLGNITVDGPLRVNNOVG 452

Qy 469 YALAGSSANFEKAGTDTKNGCTATFNNDISLGRFVNLKVDHAHTANFKG-----516
Db 469 YALAGSSANFEKAGTDTKNGCTATFNNDISLGRFVNLKVDHAHTANFKG-----516

Db 453 AAIAGSSANFEKAGEDTNNATATFNNDIHLGKAVNLRVDAHTANFNNGNIYLGKSTNLRV 512
QY 517 -----IDT--GNGFN--TLDFSGVTDKVNINKLITASTNVAVKNFNINELIVKTN 563
Db 513 NGHTAHFKNIDATKSDNGLNTSTLDFSGVTDKVNINKLITATNVAVKNFIDIKELVVITR 572
QY 564 GISVGEYTHFSDISQSQSRINVRLETGRSLRSGVGFKEGKGLVIDEFYSPMNYFDA 623
Db 573 VQSGGYTFIFGNIQDKSRIGVSVLQTYGSPAYSGGVTFKGGKGLVIDEYIHPAPNYFDA 632
QY 624 RNIKWEITNKLAFGPGSPWGTSKLMFNLLTQGNVMDYSQFNSLLTQGFINNNGTI 683
Db 633 RNVTDVEINKRLTGAAGNAGTGLMFNLLTNSASMDYDKDLDTLTQGHFTNNGTM 692
QY 684 NYLVGGKVATLSVGNAAAMNNDISATGYKPLIKINSQDLIKNTEHVLAKKIIG 743
Db 693 NLFVODGRVATLNAGHQASMIENLVDSITGYKPLIKINNAQNLTKNKEHVLVARNID 752
QY 744 YGNVSTGTNG-----ISNVLEEQFKERLALYNNNNRMDTCVVR--NTDDIKACGMAI 794
Db 753 YNLV--GVQASGYDNISASNTNLQEQFKERLALYNNNNRMDTCVVRKDLNDIKACGMAI 810
795 GQOSWVNPNPNKYLLIGRAKWKIGISKTANGSKISVIYLCNSTPTENGNTTNLPTNTS 854
Db 811 GQOSWVNPNPNKYLLIGRAKWKIGISKTANGSKISVIYLCNSTPTENGNTTNLPTNTN 870
QY 855 NARSANALQAQAPASATNLVAINOHDFGTIESVPELANRSKDIDTLVANSQAQR 914
Db 871 NARFASYALIKNAPAH--SATNLVAINOHDFGTIESVPELANRSKDIDTLVANSQAQR 929
QY 915 DLLQTLTLLIDSHDAGYARKMIDATSAEITKQLTNTATTTLLNNIASLEHKTSGLTLSNA 974
Db 930 DLLQTLTLLIDSHDAGYARKMIDATSAEITKQLTNTATTTLLNNIASLEHKTSGLTLSNA 989
QY 975 MTLNRLNLSRHHNHIDSFARLQALQDKQFASLESAAEVLVYQFAPKYERTNWNANA 1034
Db 990 MTLNRLNLSRHHNHIDSFARLQALQDKQFASLESAAEVLVYQFAPKYERTNWNANA 1049
QY 1035 IGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGFGSYGYSFNNRANSLNSGANNTNF 1094
Db 1050 IGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGFGSYGYSFNNRANSLNSGANNTNF 1109
QY 1095 GYVSIRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNOSYHYLAYSAASTRASYGYDFA 1154
Db 1110 GYVSIRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNOSYHYLAYSAASTRASYGYDFA 1169
QY 1155 FFRNALVLKPSVGVSYNHLGTFNFKSNSTNOVALKNGSSQHLFNASANVEARYYYGDT 1214
Db 1170 FFRNALVLKPSVGVSYNHLGTFNFKSNSTNOVALKNGSSQHLFNANANVEARYYYGDT 1228
1215 YFYMAGVLQEPFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGELKLAKVEFLNLGVV 1274
Db 1229 YFYMAGVLQEPFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGELKLAKVEFLNLGVV 1288
QY 1275 YLHNLSISNTHGPFASNLGMRYSF 1296
Db 1289 YLHNLSISNTHGPFASNLGMRYSF 1310
RESULT 7
OMP_RICRI STANDARD: PRT: 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Rickettsiaceae; Rickettsieae; Rickettsia.
NCBI_TaxID=783;
[1]
SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
rickettsii is encoded by an unusually long open reading frame:
evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC !- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC !- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY.
CC !- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
EMBL: X16353; CAA34403.1; --
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1191 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
Query Match 5.0%; Score 335.5; DB 1; Length 1654;
Best Local Similarity 20.5%; Pred. No. 4.2e-10;
Matches 292; Conservative 169; Mismatches 516; Indels 445; Gaps 64;
QY 138 DAVGTYTTLGSLRNFTHGGDLDDV-NMOKATLRL-----GQ----FNGN----- 173
Db 203 NAVIANGVNGTLNVTNGFTQVSNKSPATYKAINIADQGGIIFNTDANNANTLNLAGGIT 262
QY 174 -SFTSYKDSADTRTRVDFNAKNISIDNFVEINNRVGS-GAGRKASSTVTLTQASEGITS 231
Db 263 INF-----GTDGTGRVLVLLSKHAAATNF---NITSLGSLNKLGVIEFTVAVDGLTAN 314
QY 232 KNAEISLYDGLATNLASSSVKLMGNVMMGRLOYVGYAYLAPSYSTINTSKVTGEVNEHL- 290
Db 315 AGAANAVI--GTNNGAGRAAGFVSVDNKGKATIDGQVYAKDMVQSANATGOVNERHIV 372
QY 291 -----TVGDKNAQAAGIIANKKTNIGTLTLQMSAGLNIIAPP-----EGGYDKPNTPTS 340
Db 373 DVGADGTTAFKTA--SKVTITQDSNFGNTDFGNLAA-QIKVPNAITLTNFTGDSASNPGN 430
QY 341 QSGA-----KNDKNESAANDKQESSQNN----- 363
Db 431 TAGVITFDANGTLESASADANVAVTNNITAEASGAGVYVQLSGTHAAELRLNAGSIFKL 490
QY 364 SNTQVINPNSAQKTEVQPTVIDGPFAGKGTQVNVNININTNADGTIRVGG----- 415
Db 491 ADGTVIN-----GKVNQTLVGGALAAAGTITL-----DGSATITDIGNAGGA 533
QY 416 -----FKASLTNAAHLHIGKGVNLSNOASGRSLIVENLTGNITVD----- 457

[illegible]

RESULT 8

146	QY	146	SGURNETGDLDVNMOK--ATLRLGQFNGNSFTYSKDSADRTTRVDENAKNISIDNFVEIN	204
Db	211	Db	NGPLNTNGFIKIVSSKSFATVNWINDGOGIMFTDNDKDNVNTLMLQANGATI----	TFN
QY	205	NRVSGC-----AGRKASSTVLTLOAS-----EGI-----TSDKNAEISLYDGA-----	TLN	245
Db	267	Db	GTDTGRLVLLSKNAAATDFNTVSLGGNLKGIIEFNTVAVNGQLKANAGANAAVIGFNN	326
QY	246	246	LASSSVKLMGNYMQLQVGYALAPSYSTINTSKVTGEVNFHL-----TVGDKNAA	298
Db	327	GAGRAAGFVVVDNGKVAIDGOVYAKDMVIOQANAVGOVNFHRHVDVCTDGTAFKTA	386	
QY	299	QAGIIANKTNTGTLDLWOSAGLIIAPPE-----GGYKDKPNTP-----SQS	342	
Db	387	-SKVAITONSNEFTDFGNLAA-QIIVPNTWTLNGNFTGDASNPGNTAGVITFDANGT	444	
QY	343	GAKNDKNEAKNDKQESSNNSTQVIPPNSAQ-----KTEVQPTQ	384	
Db	445	SASADANVAVTNNITAIESAGVAVQLSGTHAEELRLGNAGSVFKLADGTWINGKVNQ	504	
QY	385	VIDGPAGGKDTV---VNTNRIINTNADGTIRVGGF-----KASLTWNAAHLHGKGGVN	435	
Db	505	LVGCGALAGTITLDGSATTGDI GNAGGAAALQGTITLANDATKTIITLCGANI-ICANGGT	563	
QY	436	LSNQASGRSLIVENTGNITVDPGLRVNNQVGGYALAGSSANFEE-----KAGDTKNG-	489	
Db	564	INFOANGGTIKLTSTQNNIVDFDLAIAFDQTVGVVDASSLTNAQTLTINGKIGITVGAN	623	
QY	490	-----TATPNDISLGRFV-----NLKVD	508	
Db	624	TLGQFNIGSKTVLSGDVAINELVIGNNGAVQFAHNTYLTTRTTNAAAGQGIIFNPV	683	
QY	509	AHTANFKGIDTNG-----GFNTLDFSGVTDKVNINKLITA-STNVAVKNFINELIV	560	
Db	684	NNTLATGNLSATNPPLAEINFGSGKAANDVTVLNCGKVNLYATNITTTDANVGSP	743	
QY	561	KTNGISVGYTHFSEDIGSQ--SRINTVLETGTRSLFSGGKFKGGEKLVIDEFYSPW	618	
Db	744	NAGGTNI-----VSGTVGGQGNKENTVALDNGTTVKFLGNATGNTTIAANSTLQIG	798	
QY	619	NY-----FDARNIKNVEITN-----KLAFGQGGSP-----WGSYKLMFNMLTGQ	658	
Db	799	NYTADFAVASADGTGI--VEFVATGPTITVTLNQAPAPNALKOITVSGPGNVVINEIGN	856	
QY	659	N---AVMDYSQFSNLTIOGDFINNOCTINLYVRGCKVATLSVGNAAMMENNDI--DSAT	713	
Db	857	NHGCAVTDIIAENSLS-----GAVVFLPRG--IPFNDAGNRIPIITKSTVGNKRTAT	906	
QY	714	GFYKPLIKINSAQDLKNTHEVLLKAKIIGY-----NVSTGTNGISNVNLEE---	761	
Db	907	GEDVPSPVILGVDSVIADQGVIGDQNNIVGLGSDNDIIVNATTLYAGITINNNOGTV	966	
QY	762	-----QKERLAL--YNN-----NNRMDTCVVRNTDDIKA	789	
Db	967	TLSGGIPNTPGTVYGLGTGIGASKFKQVTFDTYDNNLGNLIATNATINDGVTVTTGGIAG	1026	
QY	790	CG-----MAICDQSMVNNPNKY-----LICKANKNTGISTKANGSKISVYILGNS-----	836	
Db	1027	IGFDGKITLGS-----VNGNGNRFVDGILSHSTSMIGTTKANNG---TTVYLGNAFVG	1080	


```

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Nakano K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshino T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC [- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CVSG (AC P25928).
CC [- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC DR EMBL; AE000237; AAC74483.1; ALT_SEQ.
CC DR EMBL; AE000237; AAC74487.1; ALT_SEQ.
CC DR EMBL; D90778; BAA15009.1; ALT_SEQ.
CC DR EMBL; D90778; BAA18880.1; ALT_SEQ.
CC DR EMBL; D90779; BAA18881.1; ALT_SEQ.
CC DR EMBL; X62680; -; NOT_ANNOTATED_CDS.
CC EcoGene; EG11307; ydha.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 489 489 I -> V (IN REF. 2).
CC CONFLICT 495 495 I -> V (IN REF. 2).
CC SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.9%; Score 326; DB 1; Length 2003;
Best Local Similarity 21.0%; Pred. No. 1.7e-09;
Matches 316; Conservative 194; Mismatches 566; Indels 426; Gaps 78;

QY 50 TGTAVGTGSLWSGLK-----QAEANKTPDKPKVWRIQAGKGFNFPNKKEYDLYRS 103
DB 165 TENADGTISLQDSNGRKATINLWQIDEANNTV-----A 197
QY 104 LLSSKIDGWDWG-NAAHYWYWG-----GQNKLEVDKMDAVGYTTLGLRN--FTGDD 155
DB 198 LEGVSADGATKQWYNHNGELVITGDNATVNNNGKTTVDGKDSGTG-EINGNNGKVIQDGD 256
QY 156 LDVNMQKATRLRCQNGNSFTSKDSADR---TTRVDENAKNISIDNFEINNRVG---- 208
DB 257 LDVS-----GGHGIDITGDSATVDNKGTTWTVDPSMGIDQIDGKAIYVNNEGESTI 308
QY 209 --SGARKAS-----STVLTLQASEG-ITSDKNAEIS-LYDGTATLNL 246

```

```

DB 309 TNGGTGTQINGDDATANNNGKTTVDGKDSGTGEINGNNGKVIQDGLDVSGGHGIDITG 368
QY 247 ASSSVKLMGNVWMLQYVGYAYLAPSYSTINT-----SKVTGE-----VNFNHL 290
DB 369 DSATVDNKGTTWTDPESIGIQVDGQAVYVNNEGESAITNGGTQINGDQATANNNGKT 428
QY 291 TVGDKNAQAAGIIAN--KKTNIGTLDLQMSAGLIITAPPEGYKDKPN-----NTPSQSA 344
DB 429 TVDGKDSGTGTAGNNGKVIQDGLDV--SGGHGIDITGDSATVDNKGTTWTDPESIGI 487
QY 345 KNDKNESAKNDKQESSQNNSTQV--IN-----PPNSAQKTEVQPTQVDFGPFAGGKDTVVN 399
DB 488 QIDGQAIYVNNEGESTITNGGTGTQINGDQATANNNGKT-----TVDGKDSGTGTIAGN 541
QY 400 INRINTNADGTTRVGFKASLTNAHLHIGKGVNLSNQASGRSLIVENLGNITV--- 456
DB 542 IGIV--NLDGSLTV-----TGAH-----GVENIGDNGTVNNKGDIVVSDTSGISGLVN 588
QY 457 -----DGPLRVNNQGGY-----ALAGSSANFEFRAGTDTKNGTATFNNDISL 499
DB 589 GEGATVSNITGDVNVSNENATGFSITTSNGKVSLAGSQVGFSTGVDL--NGN---NNSVTL 644
QY 500 -----GRFVNLKVDHTANFKG---IDTNGNGGNTLDF-----SGVTDK 535
DB 645 AAKDLKVVGQKATGINVSGDANTVNITGNVLVDKDKTADNAAEYFPDPSVGINVYGSNN 704
QY 536 VNINKLIT--ASTNVAVKNFNI--NELIVKTINGISV---GEYTHFS---EDIGSQ----- 580
DB 705 VTLDGKLTVVSDSEVTSRQSLNFDGSAKTSGLVLVIGDNGTVNMMNGGLELIGEKNALADG 764
QY 581 SRINTVVR-----LETGTRSLFSGG-----KVTQHSYFNKYSTGTGNSFNNDVSSITL 927
DB 765 SVTSLRTGYSTSVIVVSGESSVYNGTTSIGEPFLGFAVIRVODKALLEIGSGATL 824
QY 609 VIDEFYSPWNYFDARNIKNVEITNKLAF-----GPGSPMGTSKLPMNLTGLQNAVY 662
DB 825 TMQDI-----DSFEHGTTRVEIQN--LGFAFVTGNTTGINSGLTSLQNGKDPAPSPV 878
QY 663 DYSQSNLTIQGDFINNQGTINYLVRGGKA-----TSLVGNAAAMFNPNDIDSATG 715
DB 879 -----LLATNGSGSATNAGTIT-----KVTQHSYFNKYSTGTGNSFNNDVSSITL 927
QY 716 YKPLIKINSAQDLIKNTEHVLKAKIIGYNGVSTGTNGISNVNLERQFERL--ALYNNNN 774
DB 928 -----VAOSNSTIINTDSGIIDL---YGRSVGMLAATADSTAEQNGKITLDSMWVDAN 977
QY 775 RMDTCVVRNTDIKA-----CGMAIGDQSMVNNPDNYKYLIGKAWKNIGISKATAN-----G 825
DB 978 --DTTAMRDIASNSAIDFGTGVGVG-----TDSYS-----GAGKNATAINQLG 1018
QY 826 SKISVYVLGNSTPTENGNTNLTPTNTTSNARSANNAQAAPPAPSPATPNLVAINQHD 885
DB 1019 GVITIYNAG-----AGMAAYGASNTVINQGTIN--LEKNYDSDSLAANTLVGMAVYE 1069
QY 886 FGT-IESFELANRSKIDITLYANSQAQGRDLLQTLIDSHDA-----GYARKMIDATSAN 940
DB 1070 HGTAINQDQGVINVTGQAFYNDG-----TGTIVNYGTICTFGVCQSGNEYNNTD 1121
QY 941 EITKQINTA-----TTTLNNNTIASLEHKYSG-----LQTLSLSNAMILNSKLVLNRSR 987
DB 1122 DFTSLIYTGDTITRSGETVTLNKSAAVTKLAGNVNVSGLTSGDQITVSSGILLENTSG 1181
QY 988 HFNHIDSPAKRLQALKDOKFASIESAAEVLVQFAPKEYKPTVNWANAIGTSLNCSNAS 1047
DB 1182 IINNL-----VKLDKGAVIKNAGV-----TNNV--DVSGILLNAGEMTA 1220
QY 1048 LYGTSGAGDAYL---NGQVEAIVGGFGSGYGS--SFNNRANSLNSGANNTNFGVYSR---- 1099
DB 1221 QITMAGADSSLVNNTGTTINKIVQACVFNNSGVTVGRMMSAGGVFNQOTDGAIMRGAAL 1280
QY 1100 ----IFANQHEFDEAOGALGDSQSSLNFKSALLQDLINQSHYLAISAATRASGYDFAFF 1156

```

Db 1281 TGTAVANN-----EGTNLGGSSGNN--TCMLEVNNS-----AFN--NRGEFILDND-- 1325
Qy 1157 RNALVLKPSVGVSYNHLGKSTNFKNSSTNOVALKNGSSSOHLFNASANYEARYY----- 1210
Db 1326 KNAVHINQS-GTLN-TGHMINSNSHGAVMNGGNGRFRINDGTIDYSAKSLVVSANNA 1383
Qy 1211 GDTSYFYMNA--GVLQBEFAH-----VGSNNAASL 1237
Db 1384 GDQNAFFWQNDQVTF-NFDHDSASAVKVTSHNFIAQNDGIMNISGTGAVAMEGDKNAQLV 1442
Qy 1238 NTFKVNAAARNPLNTHARVMGGL----KLAKEVFLNLGVV-----YLHNLISNIGHFA 1287
Db 1443 NNGTINL--GTAGTDTGMIGMQLDANATADAVIENNGTINIFANDSFASFVSLGTGVHV 1500
Qy 1288 SN 1289
Db 1501 NN 1502

RESULT 11
OMPA_RICRI
ID OMPA_RICRI STANDARD; PRT; 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rOmpA) (rOmp A).
GN OMPA.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R.
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990).
CC -I- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -I- S-LAYER WITH HEXAGONAL SYMMETRY.
CC -I- PTM: GLYCOSYLATED (PROBABLE).
CC -I- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; M31227; AAA26380.1; .
DR PIR; A41477; A41477.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.
FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 212 286 A (TYPE I).
FT REPEAT 287 358 B (TYPE II).
FT REPEAT 359 430 C (TYPE II).
FT REPEAT 431 505 D (TYPE I).
FT REPEAT 506 577 E (TYPE II).
FT REPEAT 578 652 F (TYPE I).
FT REPEAT 653 724 G (TYPE II).
FT REPEAT 725 799 H (TYPE I).
FT REPEAT 800 874 I (TYPE I).
FT REPEAT 875 949 J (TYPE I).

FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MW; A9D6464C089DF087 CRC64;

Query Match 4.7%; Score 313; DB 1; Length 2249;
Best Local Similarity 19.5%; Pred. No. 1e-08;
Matches 354; Conservative 193; Mismatches 589; Indels 676; Gaps 83;

Qy 14 LVSIALVGVLSIIPQOSHAAFFTV-----IPIAVGGIATGTAVTVGSLLSNGLK 66
Db 31 LSSSGALGVATGVIATNNAAFSNNVNNWNEITAA---GVANGTPAGGPQN--NNAFT 85
Qy 67 QAEEANKTPDKPKVWR-----IQACKGF--- 90
Db 86 YGDYTVTADAADRIIRKAINVAGTTPVGLNITQNTVVGSIITKGNLLPVLNAGKSLTN 145
Qy 91 -----NEFPNKEYDLYRSL-----LSSKIDGMDWG--- 116
Db 146 GNNVAANHGFDPADPNYTGGLNIALGGANNAALIIQSAAPSKITLAGNIDGG---GIITV 202
Qy 117 --NAARHYWYKG--GQONKLEVDKMDAVGYTLG----- 147
Db 203 KTDAA---INGTIGTNTAL-ATVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAV 257
Qy 148 ----LRNFTGGD---LDVN--MOKATRLRLGCPNGNSFTSYKDSA----- 182
Db 258 LTGAIDNTTGGDNVGVNLNGALSQVTDIG--NTNSLATISVGAGTATLGGAVIKATTT 315
Qy 183 ---DRTTRVDFNAKNI---SIDNFEVNNRV-----GSCAGRKASSTVLT---L 222
Db 316 KLTDASASAVKFTNPVVVGTGAIDNTGNANGGIVTFTGNSTVTVGNVNTNALATVNVGAGLL 375
Qy 223 QASEGITSDKNAEISLYDGTAL-----NLASSVKLMGNVMMGRLOYVG----- 266
Db 376 QVQGGVV--KANTINLTDNASAVFTTNPVVVGTGAIDNTGNANGGIVTFTGNSTVTDIGN 433
Qy 267 -----AYLAPSYSTINTSKVTGEVFNPHLTVGDKNAOAGIATANK--FNIG 311
Db 434 TNALATVNVGAGTATLGGAVIKATTTKLTNAASV--LTLTNANAVLTGAIDNTTGGDNV 491
Qy 312 TLDLWQSGAGLNIIAPPEGYKDKNPNTPSOS-----GAKNDKNESAKNDKQESS 360
Db 492 VLNL--NGALSQVTDIG-----NTNSLATISVGAGTATLGGAVIKATTTKLTDAASA 542
Qy 361 QNNSNTQVIPPNSAQKTEVQPTQVIG-----PFAGGKDTVVNINRINTNADGTI 411
Db 543 VKETNPVVV-----TGAIDNTGNANGGIVTFTGNSTVTDIG--NTNSLATI 587
Qy 412 RVCGFKASL-----TTNAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGPL 460
Db 588 SVGAGTATLGGAVIKATTTKLTNAASV-----LTLTNANAVLTGAIDNTTGGDNV--GVL 640
Qy 461 RVN---NOV--GGYALAGSSANFEFKAGTDTKNG-----TATFNNDISLGRFV----- 503
Db 641 NLNGALSQVTDIGTNTSLATISVGAGTATLGGAVIKATTTKLTNAASVAKFTNPVVVVG 700
Qy 504 -----NLKVDAHTANFKGIDTNGGFGNTLDFSGVTDKVNINKLITAST 546
Db 701 AIDSTGNANGGIVTFTGNSTVTDIGTNTNALATVNVGAGTATLGGAVIKATTTKLTNAAS 760
Qy 547 NVAVKFN--INELIVKTNG---ISV---GEYTHSFEDIGSOSRINTVLETGTRSLFS 597
Db 761 VLTLTNANAVLTGAIDNTTGGDNVGVNLNGALSQVTDIGTNTSLATISVGAGTATL-- 818
Qy 598 GGVKFK-----GGEKLVIDEFYYSPTWTFDARNIK 627
Db 819 GGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNVGVNLN-----NGALSQV 873
Qy 628 NVEITNKLAFGPGSGPWGTSKL-----MFNNLTJL-GQNAVMDYQFSNLTJ 672

Db 44 GNTDFGNLA-----AQIKVPNAI---TLTG--NFTG---DASNPGNTAGVITF 83
QY 171 NGSFTSYKDSADRTTRVDFNAKINISIDNFVEINNRVSGAGRKGASSTVLTLQASEGITS 230
Db 84 DANG-TLESASAD-----ANVATNITATEASGAGVVQLSGT-----HAEELRLG 128
QY 231 DKNAEISLYDGATLNLASSVYKMGNYMGRLOYGVYGLAPSYSTINTSKVTGEVFNHL 290
Db 129 NAGSIFKLADGTVINGKVNQALVG-----GALAAGTITLDGSAITIGDIG--- 174
QY 291 TVGDKNAQAAGIANKNTNIGTLDLWOSAGINIIAPPEGGKDKPNPTPQSGAKNDKNE 350
Db 175 NAGGAALQRITTLANDAKTLTL-----GGANII-----GAGGCTIDLQA 214
QY 351 SAKNDKQESSONSNTQVINPPNSAQTEVQPTQIDGPFAGGKDTVVNINRINTNADGT 410
Db 215 NGGTIKLITSTQNNI-----VDFDLA-----IATDQTVG 243
QY 411 IRVGGFKASITTTNAHLHI-GK-GGVNLSNOAGRSILVENLTGNITVVDGLRNVNQVGG 468
Db 244 V-----DASSLTNAQTLTNGKIGTIGCANNTLQGFNIGSSKT--VLSNGNVAINELVIG 296
QY 469 YALAGSSANFEFRAGTDTKNGTA-----TFNNDISLGRFVNLKYDAHTANFKGIDTG 520
Db 297 NDGAQVFAHDYLI-TRTTNAAGGKIIIFNPVYVNGTTLAAGTNL--GSATNPPLAEINFG 353
QY 521 NGGFNTLDFSGVTDKVNINKLITASTNVAVKFNINELIVKTNGISVGEVTFHSEDIGSQ 580
Db 354 SKGVNVDTVLNVEGVNL-----YATNITTTDANGVGFVFNAGGTNI-----VSGTVGGQ 403
QY 581 --SRINTVRLETGRSLFSGGVKFKGKELVIDEYFYSPWNY-----FDARNIKNVEI 631
Db 404 QGNKFTVALENGTTFVAFGLNATFNNGNTTAAANSTLQIGNYTAADCVASADGTGI--VEF 461
QY 632 TN----KLAFGPGQSP-----WGTSKLMFNNTLQGN--AVMDYQSFNSLTIGD 675
Db 462 VNTGPITVTLKNAQPNVNAKQITVSGPGNVVINEIGNAGNHGAVTDITAFENSSL--- 518
QY 676 FINNOGTINYLVRGGKATVLSVGNAAAMFNNDIDSAT--GFYKPLIKINSADOLIKNTE 733
Db 519 -----GAVVELPRG--IPFENDAGNTMPLTIKSTVGNKTAKGFDVPSVVLGVDSVIADGQ 571
QY 734 HVLKAKIIGY-----NVSTGTNGISNVNLEB-----VSGTVGGTGTGTGI 631
Db 572 VIGDQNNIVGIGLSDNGIIVNATTLYAGISTLNNNGTTLTSGGVPTNPTGTVYGLGTGI 631
QY 762 ---QFKERLAL--YNN-----NNRMDTCVVRVTDIDKACG---NAIGDQSVNVPND 805
Db 632 GASKFKQVTTTDDYNNLGNIIATNATDGVTVTTGGIAGIGFDGKITLGS---VNGNGN 688
QY 806 YKY---LIGKAWKNIGIKSTANGSKISVYVLGNSTPTPENGNTNLP-----TNTTSNARS 858
Db 689 VRADGILNSTNIGTWTKANNG--TVTVLGNFVGNIGSDTTPVASVFTGSDSAGL 745
QY 859 ANNALQAQAPFAPSATPNLVAHQHDF-----GT----- 888
Db 746 QGNIYSQVIDF---GTYNLGIIVNSIILGGGTTAINGKIDLVNTLTFTFASGTSTWGNNT 801
QY 889 -IESVFELANRS-----KIDITLYAN-SGAQGRDLLQ----- 918
Db 802 SIETTLTANGNHGHIIVLEGAQVNTTTTGTIKVODNANANFSGTQVTVTLIQGGARFN 861
QY 919 -TLIDSHDAGYARKMIDAT---SANE---ITKQNTATTTLNNIASLEH----- 961
Db 862 GTLGSPPFAVTSNRFVNSLIRAAQNDYVITRTNNAENVVNTDIANSFPFGAGPVGDQNV 921
QY 962 -----KTSGLQTL-----SLNAMILNSRLVNLRRHTNHIDSFARKLOALKDKOFAS 1009
Db 922 TTFVNATNTAAYNLLAKNSANSANFVGAIVTDTSAATNVQLDLAKDIOAQLGNRLGA 981
QY 1010 LESAAEVLQYFAPKYEKPTVWNAI-----GTSLSNN----- 1042
Db 982 LR-----YLGCTPETAEMAGPEAGISAAVAAGDEAIDNVAYGIWAKPFTYDAHQSKKG 1035

QY 1043 --GSNASLYGTSAGDAYLNGQVEAIVGGFSGSYGSSFNNSRANSLNSG----ANNTFGV 1096
Db 1036 LAGVKATTTGVVIGLDTLANDNL--MIG--AAIGITKTDIKHODYKKGDKTDVNGFSESL 1091
QY 1097 YSRIFANQHEFDEFAQG-ALGSDOSSLNFKSALLQDLNQSY-----HYLAISAAATRAS 1148
Db 1092 YG---AQOLVKNFFPAQGSATFSLNQVKNKSORYPFDANGNMSKQIAAGHYDNMTFEGNLT 1148
QY 1149 YGDFAFERANLVLPKPSGVSYVSNHLGSTNFKSNSTNOVALKNGSS 1193
Db 1149 VGVDYNAQGVLV-TPMAGLSYLKSSDENYKETGTT-VANKQVNS 1191
RESULT 13
ID OMPA_RICCN STANDARD; PRT: 2021 AA.
AC Q52657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;
AC Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
antigen) (rOmpA) (rOmp A).
GN OMPA OR RC1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.-E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
analysis of restriction fragment length polymorphism of PCR-amplified
DNA of the gene encoding the protein rOmpA";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.-E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; U01028; AAL17405.1; -
CC DR EMBL; AE008674; AAL03811.1; -
DR EMBL; U43794; AAB49549.1; -
DR EMBL; U43798; AAB49550.1; -
DR EMBL; U43806; AAB49551.1; -
DR EMBL; U45244; AAB49566.1; -
DR EMBL; U46918; AAB46663.1; -
DR EMBL; U83440; AAC35176.1; -
DR EMBL; U83443; AAC35179.1; -
DR EMBL; U83448; AAC35184.1; -
DR EMBL; U83453; AAC35189.1; -
DR InterPro; IPR003858; rOmpA_rOmpB.
DR pfam; PF02708; rOmpA_rOmpB.1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1877 1877
FT CONFLICT 10 10
FT CONFLICT 92 92
FT CONFLICT 126 126
FT CONFLICT 137 137
FT CONFLICT 157 157
FT CONFLICT 368 369
FT CONFLICT 374 388
FT CONFLICT 640 640
FT CONFLICT 669 669
FT CONFLICT 793 793
FT CONFLICT 803 804
FT CONFLICT 809 823
FT CONFLICT 898 898
FT CONFLICT 908 908
FT CONFLICT 985 985
FT CONFLICT 1009 1009
FT CONFLICT 1013 1013
FT CONFLICT 1182 1182
FT CONFLICT 1314 1314
FT CONFLICT 1451 1451
FT CONFLICT 1624 1624
FT CONFLICT 1628 1628
FT CONFLICT 1872 1872
FT CONFLICT 1875 1875
FT CONFLICT 1878 1879
FT CONFLICT 1936 1936
FT CONFLICT 1965 1970
FT CONFLICT 1997 1997
SQ SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;

Query Match
Best Local Similarity 4.7%; Score 310.5; DB 1; Length 2021;
Matches 302; Conservative 207; Mismatches 533; Indels 463; Gaps 80;

QY 19 LVGALV-----SITPQSHAAFFTTVIIPAIV-----GGIATGTAAGTAVGTVSGLLSWGLK 66
DB 740 LGGAVIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIATFTGDSVTGNTG---- 795
QY 67 QAEEANKTPDKPKVVRIOAQ-----KGFNEFPNKEYDLYRS--LLSSKIDGWDGNAAR 120

Db 796 --NTNALATVNVGAGLLRVGGVVKSGNTINLTNDASAVTFTNPVVVTGAID---NTGNANN 851
QY 121 HWYKGGQONKLEVDKMDKDAVGYTTLGSLRNFTE-----GDLQVNMOKATLRLQGFNG 172
Db 852 -----GIVTFTGDSVTGNTGNALATISVGAGKATLG-GAIK 890
QY 173 NSFTSYKDSADTRTRVDFNAKNISIDNFEVNNRVSGAGRKASSTVLTLQASEGITS DK 232
Db 891 ATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIIVTGT---DSTV-----TGNI 941
QY 233 NAEISLYDGCATLNASSSVKLMGNVMMGRLOV-----GAYLAPSY----- 273
Db 942 NALATVNVGAGVTLQAGGSLDANNIDFGARSTLEFNGPLDGGGAIPYFYKGAIANGNA 1001
QY 274 -STINTSKVTGEVNFENHLLTVGDK-----NAAQAGI 302
Db 1002 ILNVTKLLTA-----YHITGTVAEINIGAGNLFDAIDASAGDVTILNQADIIHFRALDSAL 1057
QY 303 TANKKTNICTLDLQMSAGLIITAP--PEGYKDKPNNTPSQSGAKNDKN-----ESAKND 355
Db 1058 VLSNLTGCVGNVILLAA--DLVAPGVDEG-----TVVFDGGVNLGNSVAGAAARN 1107
QY 356 KOESSQNSNTQVNIAPPNSAQKTEVQPTQVIDGPPAGGKDTVVNNIRINNTAD----- 408
Db 1108 IGDVGKGNKFNLLI-----YNAVITDDVNLEGIQNVL-----INNADFTSSSTAF 1153
QY 409 --GTRVGGFKASLTNAHLHIGKGVNLSNQASGRSLIVENLTGN---ITVDGPLRVN 463
Db 1154 NAGTIQINDATYTTDANNGNLNPAGNIKFAH--ADAQLIOLSSGNDRTITLGANIDPD 1211
QY 464 NO-----VGYALAGSS--ANFEFRAGTDTKNGTATFN-NDISLGRF 502
Db 1212 NDDEGIVILNSVTAGKLTAGGKTFGGAHKLQDIVRKEGDFGTAGTFTNTNIVLDT 1271
QY 503 VNLKVDATAN---FKG-----IDTGN-GGFNTLDFSGVTDKVNINKLITASTNVA--VKN 552
Db 1272 GQLELGATTANVLEFKDAVQLTQTGNIIGF--LDFNAKNGVTYTLNN---NYNVAGTVKN 1325
QY 553 ---FNINELIV-----KNGISV-----GRYTHFSEDIGSOSRINTVRLTGTSL-- 595
Db 1326 TCGTNGTGLVLGASNLNRVNGIAMLKVAGNWTAK---GCVKIGEIQ-GTGNLTLL 1381
QY 596 -----FSGGVKFKGKGLVIDEYFSPWNYFD-----ARNIKNVEITNKLAQFP 639
Db 1382 PAHEKLTGSINKTGOALKL-----NEMNGSVGVVGTAAVSGDITTAGATSPAS 1433
QY 640 QGSPWGTSKL-----MFNN---LTGQNAVMDYSQFSNLTIOGDFINNOGTINYLVR 688
Db 1434 SVNAKGTATLGTTSTFAHTFTNTGAVTLAKGSITSFAK--NVTAT-SFVANSATINF--- 1487
QY 689 GGVKATLSVGNAAAMFNNDID-----SATGFYKPLIKINSAQD-LIKNT 732
Db 1488 -----GNSLA--FNSNTGSGTTLTGANQVYTYGTGSGTDTTLNTTFDGAAGSG 1536
QY 733 EHVLLKAIIGYGNVSTGTNGISLVNLEEQFKERLALYNNNNRMBDTCVVVRNTDDIKACGM 792
Db 1537 GNILIKS-----GSTLDLSGVSN-----LALVVTATNFD----- 1565
QY 793 AIGDOSMVNN--PDNYKYLIGKAWNIGISKATANGSKISVYVLGNSTPTENGNTNLTPT 850
Db 1566 -----MNNISPDIT-KYTVISA-----ETAGGLK-----PTPKENYKTIINDN 1602
QY 851 NTSNARSANALAQNAFPAQSPATPNLVAIINHDF---GTIESFELANRDKDITLY- 906
Db 1603 RVDVTFDASTLTL-----FAEDIA-----AGVIDEFAQGGPIANIPNANIKKSLELMD 1654
QY 907 ANSQAQGRDLLOT--LLDSDHDAGYARKMIDATS--ANEITQLNTATTTLNINSLAEHK 962
Db 1655 APNGSDARQAFNFGMLTLPLOEA-----DATHLMQDVVKPSTIAAVNNQVVA--- 1703
QY 963 TSGLOTLSLNAIINSLRLVNSRHTNHIDSFARKLQALKDQKFALESAAAEVLYQFAP 1022

```
Db 1704 ----SNIS-SNITARNMDKVQAGNKPVSSEDMD-----DAKFGA-----WISP 1745
Qy 1023 KYEKPNTVWANAIGGTSNLNGSNASLXCTSGAVDAYLNGQVEAIVGGFGSYGSF----- 1078
Db 1746 FVGNAQTQMCNSI-----SGYKSDTTGGTIGDPGFSVD--DLVLG-----LAYTRADTDI 1793
Qy 1079 ---NNRANSLUNGANN--TFNGVYSRIFAN-----QHEDFPAQAGALGSDQSSLN 1123
Db 1794 KLNKNTGDKNKVESNIYSLGLYSVPYENLFVEAIIASYSNDKIRSKSRRIATTLLETVG 1853
Qy 1124 FKSALLQDLQSYHYLAISAATRASGYDEFAFRNALVLKPSVCVSYNHLGSTNFKSNST 1183
Db 1854 YQTA-----NGKYKSESYTGOLMAGYTY---MMSENINLPLAGLRVSTIKDKSYKETGT 1905
Qy 1184 NOVAL-----KNGSSQHLFNA--SANVEAR-----YYYGDTSYFYMNAGVLOEFAHVG 1231
Db 1906 TYQNLTVKGNKYNTFDGLGAKYSSNINVAEIVLTPELYAMVDYAFKN----- 1953
1322 NNAASLNTFKVNAARNPLNTHARVMMGGEKLKAEVF-LNKGVVYLLNHLISNIGHFASNL 1290
Db 1954 -----KVSADARLOGMTAPLPTNSFKQSKTSPDVGVGVTAKHKMMEYGINYDTNI 2004
Qy 1291 GMY 1294
Db 2005 GSKY 2008

RESULT 14
AIDA_ECOLI
ID AIDA_ECOLI STANDARD; PRT; 1286 AA.
AC Q3135;
DT 01-JUN-1994 (Rel. 29, Created)
DI 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-I precursor.
GN AIDA-I.
OS Escherichia coli.
OG Plasmid pIB6.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
"IDA-I, the adhesin involved in diffuse adherence of the
diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
synthesized via a precursor molecule.";
RL Mol. Microbiol. 6:1539-1546(1992).
CC -!- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65022; CAA46156.1; -.
CC PIR; S28634; S28634.
CC Cell adhesion; Signal; Outer membrane; Plasmid.
KW SIGNAL
FT CHAIN 1 49
FT PROPEP 50 ? ADHESIN AIDA-I.
FT PROPEP ? 1286
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;
```

Query Match

4.6%; Score 310; DB 1; Length 1286;

```
Best Local Similarity 21.1%; Pred. No. 6.5e-09;
Matches 304; Conservative 196; Mismatches 568; Indels 376; Gaps 71;
Qy 38 TVIIPATVGGIATGAV---GTVS--GLLSWGLKQ-----AEEANKTPDKPKVMRIQAG 87
Db 34 TLLVLAVVSTIGNAFANVIGTSSGGTSSGGETQIYVSGRGSNATVNSGGTQIVVNGG 93
Qy 88 K----GFNEFPNKEYLDLYRSLLSKIDGWM-----DMGNAARHYVWVGQONKLEVDKDA 139
Db 94 KTTATTVNSGSSQNVGTFSGATISTIVNSGGIQRVSSGGVASATNLSCGAQN--IYNLGA 151
Qy 140 VGYITLSGLRN--FTGGDL--VNMQKATRLGQPNGNSFTSYKDSADRTTRVDFNARNIS 196
Db 152 SNTVIFSGGQTIFSGGTTDSTNISS-----GGQORVSSGGVASNTTINSGAQNI- 202
Qy 197 IDNFVEINNRVGGG-----AGRKASSTVLTQASEGITSKNAEISLYDGATLINLASS 250
Db 203 LSEBGAISTHISISSGQNYISAGANATETIVNSGGFQRVNSGAVATGTVLSGGTQNVSSG 262
Qy 251 VKLMGNVMMGRLOYGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQAQAIITANKTNI 310
Db 263 SAISTSVYNSGVQTVFAGATVTDITVNSG-----GNQNISSGGIIVSETTVNV 309
Qy 311 -GTLDLWQSAGLNIIPPEGGYKDKPNPTFSQCAKNKDNESAKNDKQESQNNSTQVI 369
Db 310 SGTQNIY-----SGG-----SALSANIKGSQIV 332
Qy 370 NPPNSAQKTEVQP---TQVIDGPFAGGKDTVWN-----INRIN-----TNADGTRV- 413
Db 333 NSEGTAINTLVSDGGYQHIRNGGIASG--TIVNOSGVYVNISSGGYAEISTINSGGTURLV 390
Qy 414 --GGFKASLTNAAHLHIGKGVNLSNQASGRSLIVENLTGN--ITVDGRLRVNNVGGY 469
Db 391 SDGVARGTILNNSGRENVSNGVSYN-----AMINTGGNOVIYSDGEATA----- 435
Qy 470 ALAGSSANFEFKAGTDTKNGTATFNNDISLGRFVNLVKDAHTANFKGIDTNGGNTL-- 527
Db 436 AIVNTSGFORINSQ-----GTAPVQNSVVVTRTVSSAAKPFDA-----EYVSGKQTVYL 485
Qy 528 -----DFSGVTDKVNIN-----KLITASTNVA 549
Db 486 WRGIWYSNFLTAVWSMPPGTASGANVNLGRNLNFAFAGNVVGTILNQGRQVYVSGATATS 545
Qy 550 VKNFNINELIVKNGISVGEYTHFS--EDIGQSQRIN-TVRLETGTRSLFSG-----GVKF 602
Db 546 TVGNNEGREYVLGGITDGTIVNSGGIQAQVSSGKASATVINEGGAQFVYDGGQVGTGNI 605
Qy 603 KGEKLVIDE-----FYSPWNYFDARNIKVVEITNKLAFGPGQSPWGTSKLMPNLT 655
Db 606 KNGGTIRVDSGASALNIALSSGNNLTSTGATLPELTMAAL-----SVSQNHASNI 658
Qy 656 LGQNAVMDYSQFNLTIQGDFINNOGTFINVLVRGKVKATLSVGNAAAMFNNDIDSATGF 715
Db 659 LENGGLLRVT--SGGTATDTVNSAGRLR-IDDGGTNGTNGTTINADGIVAGTINQDGNF 715
Qy 716 YKPLIKINSADLIKNTPE-----HVLK--AKIIGYVNVSTGTNGISNVNLEOPKRLAL 769
Db 716 I-----LNLAEYDFETELSGSGVLVKDNTGIMTYAGTLTQAQGVN-----V 757
Qy 770 YNNNRMDTCVVRNTDDIKACGHAIGDSQSVNPNPDNYKYLIGKAKWKGIGISKTAGNSKIS 829
Db 758 KNGGIIFDSAVV-NAD-----MAVQNQAYINTSD--QATINGSVNNNG-SIVINNSIIN 807
Qy 830 VYVLGNST-----PTENG--NTTNLPNTTNSARSANALQAQAPPAQS 873
Db 808 ----GNTNDADLSFGTAKLLSATVNSGLVNNKNIILNPTK--ESAGNTTLVSNYTTGPG 861
Qy 874 ATPNLVAINHDFGTIESVPELANRSKIDITLYAN-----SGAQGRDLLQTLIDSH-DAGY 929
Db 862 SVISLGGVLEGDNLSLTDRLVVKNGTSGSDIVVYVNEGSGGQTRDGINIISVEGNSDAEF 921
Qy 930 ARKMIDTASNEITKQNLNTATTATTLNNTASLEHKTSGLOT-----LSLSN 973
```

Db 922 SLKNRVAGAYDTTLQKNGESGTDKNGWYL---TSHLPTSDTRQYRPENGSVATNMALAN 978
QY 974 AMILNSRLVNLSSRRHTNHDSPAKRLQALKDQKFALESAAEVLVQFAPKYEKPTNVAN 1033
Db 979 SLF-----LMDLNR-----KQFRASDN--TQPSA-----SVWMK 1008
QY 1034 AIGGTS-----LNGSN---ASLYGTSGVDAY-----LNGQVEAIVGGFGSYGSFNNR 1081
Db 1009 ITGGISSGKLDNQKTTTQINQGLGGDIYKFAEQLGDFTLGIMGGVANAAGKXTINTV 1068
QY 1082 ANSLNSGANNT-----NFGVYSRIFANQHEFDEAGALGSDOSSLN--PKSALLQDLNOSY 1136
Db 1069 S---NKAARNTLDGYSVGYGTWYNGE-----NATGLFAETWMOYNWFNASVKGDLLEE 1121
QY 1137 HYLAYSAASTRASGYDEAFERNAVLKPSVGVS-----YNHLGS-----TNFKSNST 1183
Db 1122 KYNLGLTFASAGGGYNL-----NVHTWTSPGITGEFWLQPHQAVVMGVTPDTHQEDNGT 1177
QY 1184 -NQVALKNGSSSHLFENASAEVARYYYGDTSYF--YMNAGVL---QEFVHVGSSNAASL 1237
Db 1178 VVOGAGKNIQTKAGIRASWKVKYKSLDKDGTGRFRPYIEANWIHNTHEEGVKMSDQSOLL 1237
1238 NTFKVNARNPLNTHARVMMGELKAEVFLNGLVYVYLNHLNISIGHFASN-----LGM 1292
Db 1238 -----SGSRN-----QCEIKTGIEGVITONLSVNGGVAY-----QAGGHSNAISGALGI 1282
QY 1293 RYSF 1296
Db 1283 KYSF 1286

RESULT 15
OMP_RICPR
ID OMP_RICPR STANDARD; PRT: 1643 AA.
AC Q53020: Q53020;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOmpB)
DE (Omp B) [Contains: 120 kDa outer membrane protein (Surface protein
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
DE OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
SC STRAIN=BREINL;
MEDLINE=91045972; PubMed=2122457;
Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
"Characterization of the gene encoding the protective paracrystalline-
surface-layer protein of Rickettsia prowazekii: presence of a
truncated identical homolog in Rickettsia typhi.";
Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
"Sequence analysis of ompB of Rickettsia prowazekii.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E.
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria.";
RL Nature 396:133-140(1998).
RN [4]

PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
the S-layer protein antigens of Rickettsia typhi and Rickettsia
prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
[5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
membrane protein of rickettsiae: identification of an avirulent mutant
deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC
DR EMBL; M37647; AAA26390.1; ALT_INIT.
DR EMBL; AF161079; AAD42234.1; -.
DR EMBL; AJ235273; CAAL5140.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT CONFLICT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 TTQAPLTLGA -> INSRSSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392B6346CC CRC64;

Query Match 4.6%; Score 306; DB 1; Length 1643;
Best Local Similarity 18.6%; Pred. No. 1.5e-08;
Matches 293; Conservative 179; Mismatches 556; Indels 544; Gaps 67;

QY 20 VQALVSIPTQOSHAAFFTVIIPAIVGGIATGTAVTYSGLLSWGLKQAEANKPDKPD 79
Db 109 LGFITNIAQQ-----AKFFNFTV-----AAGKILNITGQGI-TVQEAASNTINAO 152
QY 80 KVMRIQAGKGFN-----EF-----PNKEYDLVRSLLSSKIDGGMDW 115
Db 153 ALTKVHGGAANANDLSGLGSITFAAAPSVLEFNINPTQEAPLTLGANSKIVNG--- 209
QY 116 GNAARHYWVGQKQKLEVDMDKDAVGTVTLGSLRNFPTGGDLVDNMOK-----ATLRLGQFN 171
Db 210 -----NGLNITNGFIQVSDNTFAGIKTINIDCQ 239
QY 172 GNSFTSYKSDADRTTRVDENAKNISIDNFVEINNRVCSGAGRKASVTVLTQASEGITS 230
Db 240 GLMFNSTPDAA-----NTLNQVGGNTINFNGIDGTGK-----LVLSVNGAAT 283
QY 231 DKNABISLYDG-----ATLNASSSVKLMG-----NVWMGRILQ 263

Search completed: August 16, 2002, 11:34:26
Job time: 337 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:28:34 ; Search time 46.36 Seconds
(without alignments)
4836.096 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQPHRKINRPLVSLVY.....HNLISNIGHFASNLGMRYSF 1296

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6353.5	95.2	1295	2	034110 helicobacte
2	6273	94.0	1290	2	034111 helicobacte
3	6246	93.6	1290	2	09ZHT1 helicobacte
4	6180	92.6	1296	2	09ZHT2 helicobacte
5	6165	92.4	1296	2	09ZHT3 helicobacte
6	6163	92.4	1296	2	09ZHT4 helicobacte
7	6162	92.4	1288	2	09ZHT5 helicobacte
8	6154	92.3	1296	2	09ZHT6 helicobacte
9	6135.5	92.0	1291	2	09ZHT7 helicobacte
10	6128.5	91.9	1291	2	09ZHT8 helicobacte
11	6122	91.8	1296	2	09ZHT9 helicobacte
12	6120.5	91.7	1291	2	09ZHT10 helicobacte
13	6117.5	91.7	1291	2	09ZHT11 helicobacte
14	6113	91.6	1290	2	09ZHT12 helicobacte
15	6112.5	91.6	1291	2	09ZHT13 helicobacte
16	6112.5	91.6	1291	2	09ZHT14 helicobacte

17	6111.5	91.6	1289	2	093UM6 helicobacte
18	6105.5	91.5	1291	2	09ZHU9 helicobacte
19	6105.5	91.5	1291	2	09ZHT2 helicobacte
20	6101.5	91.5	1291	2	09ZHV3 helicobacte
21	6100.5	91.4	1287	2	093UM9 helicobacte
22	6089.5	91.3	1291	2	09ZHU7 helicobacte
23	6081.5	91.2	1291	2	09ZHU4 helicobacte
24	6080.5	91.1	1291	2	09ZHU6 helicobacte
25	6075.5	91.1	1291	2	09R961 helicobacte
26	6069.5	91.0	1300	2	09ZHU0 helicobacte
27	6064.5	90.9	1291	2	09ZHT9 helicobacte
28	6064.5	90.9	1293	2	09ZHV0 helicobacte
29	6061	90.9	1298	2	09ZHT4 helicobacte
30	6060.5	90.8	1291	2	09R962 helicobacte
31	6033.5	90.4	1287	2	093UM7 helicobacte
32	6012	90.1	1294	2	09R959 helicobacte
33	5862.5	87.9	1303	2	09KJA6 helicobacte
34	5574.5	83.6	1323	2	087018 helicobacte
35	5560	83.3	1324	2	09LBC3 helicobacte
36	5552	83.2	1324	2	09LBC2 helicobacte
37	5539	83.0	1324	2	09LBC7 helicobacte
38	5520	82.7	1328	2	09LBC8 helicobacte
39	5181.5	77.7	1313	2	093UN0 helicobacte
40	5114.5	76.7	1082	2	09R964 helicobacte
41	3929.5	58.9	829	2	09F9F8 helicobacte
42	3926.5	58.9	829	2	09F9G0 helicobacte
43	3918	58.7	826	2	09F9G4 helicobacte
44	3683.5	55.2	839	2	093Q01 helicobacte
45	3434	51.5	866	2	09F9G6 helicobacte

ALIGNMENTS

RESULT 1

ID	034110	PRELIMINARY;	PRT; 1295 AA.
AC	034110;		
DT	01-JAN-1998	(Tremblrel. 05, Created)	
DT	01-MAY-1999	(Tremblrel. 10, Last sequence update)	
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)	
DE	VACUOLATING CYTOTOXIN.		
GN	VACA.		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;		
OC	Helicobacter		
OX	NCBI_TaxID=210;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=F37;		
RX	MEDLINE=97339580; PubMed=9196179;		
RA	Ito Y., Azuma T., Ito S., Miyaji H., Hirai M., Yamazaki Y., Sato F.,		
RA	Kato T., Kohli Y., Kuriyama M.;		
RT	"Analysis and typing of the vacA gene from cagA-positive strains of		
RT	Helicobacter pylori isolated in Japan."		
RL	J. Clin. Microbiol. 35:1710-1714(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=F37;		
RX	MEDLINE=98453456; PubMed=9780260;		
RA	Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,		
RA	Kuriyama M.;		
RT	"Full-length sequence analysis of the vacA gene from cytotoxic and		
RT	noncytotoxic Helicobacter pylori."		
RL	J. Infect. Dis. 178:1391-1398(1998).		
RN	[3]		
RP	SEQUENCE OF 638-752 FROM N.A.		
RC	STRAIN=JK2-55;		
RA	Yamaoka Y., Kodama T., Graham D.Y., Kita M., Imanishi J., Kashima K.;		
RT	"vacA genotypes of Helicobacter pylori in relation to cagA status,		
RT	cytotoxin production, or clinical outcome."		
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF071095; AAC77450.1; -.		

Qy	1	MEIQOTHRKNRPLVSLALVGLVSTIPQOSHAAEFFTVIIPAIIVGGIATGATVAGTVSGL	60
Db	1	MEIQOTHRKNRPLVSLALVGLVSTIPQOSHAAEFFTVIIPAIIVGGIATGATVAGTVSGL	60
Qy	61	LSWGLQKBEANKTPDKPKVWRIQAGKGFNEPKNKEYDYLRSLSSKIDGGWDGNAAR	120
Db	61	LSWGLQKBEANKTPDKPKVWRIQAGKGFNEPKNKEYDYLRSLSSKIDGGWDGNAAR	120
Qy	121	HYWYKGQGNKLEVDMKDVGTVTSLGRLNFPGDLDVNMOKATLRLQCFNGNSFTSYKD	180
Db	121	HYWYKGQGNKLEVDMKDVGTVTSLGRLNFPGDLDVNMOKATLRLQCFNGNSFTSYKD	180
Qy	181	SADTRTRVDNNAKNI SIDNFEINNRVGGAGRKASSTVFLIQASEGITSDKNAEISLYD	240
Db	181	SADTRTRVFNNAKNI SIDNFEINNRVGGAGRKASSTVFLIQASEGITSGKNAEISLYD	240
Qy	241	GATUNLASSVKLGMVNMGRLOQYVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQ	300
Db	241	GATUNLASNSVKLGMVNLGRLOQYVGYLAPSYSTINTSKVVGVEVNFNHLTVGDORNAQA	300
Qy	301	GIATKNTNICTGLDQWAGSLNI IAPPEGGYKDKPNNTPSQSGAKNDKNESAKNDQESS	360
Db	301	GIATKNTYIGTLDLQWAGSLNI IAPPEGGYKDKPNNTSOSVGSKDNESAKNDKQD--	358
Qy	361	QNNSTQVINPPNSAKTEVQPTQVJIDGPFAGGKQTVVNIINRINTNADGTTIRVGGFKASL	420
Db	359	--SNTQVINPPNSGGKTEIQTQVJIDGPFAGAKQTVVNIINRINTNADGTTIRVGGYTASL	415
Qy	421	TTNAAHLIHCKGVNLSNQASGRSLIVENLTGNITVDGPLRVNNQVGYALAGSSANFEP	480
Db	416	TTNAAHLIHCKGVNLSNQASGRSLIVENLTGNITVDGPLRVNNQVGYALAGSSANFEP	475
Qy	481	KAGTDTKNGTATFNNDISILGRFVNLIKVDHAFTANFKGIDTNGGNTLDFSGVTDKVNINK	540
Db	476	KAGTDTKNGTATFNNDISILRRFVNLIKVDHAFTANFKGIDTNGGNTLDFSGVTCVKVNINK	535
Qy	541	LITASTNVAVKFNINELIKVTNGISVGEYTHFSEDIGQSRIINTVRLTGTSLFSGGV	600
Db	536	LITASTNVAVKFNINELIKVTNGVSGEYTHFSEDIGQSRIINTVRLTGTSTRSIYSGV	595
Qy	601	KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQSPMGTSKLMFNNTLIGQNA	660
Db	596	KFKGGEKLVINDFYAPWNYFDARNIKNVEITNKLAFGPQSPMGTSKLMFNNTLIGPNA	655
Qy	661	VMDYSQFSNLTIOGDFINNOGINTYLVRGGKVATLSVGNAAAMFNNDIDSATGYFKPLI	720
Db	656	VMDYSQFSNVTIOGDFINNOGINTYLVRGGKVATLNVGNAAAMFNNDIDSATGYFKPLI	715
Qy	721	KINSAQDLIKNTEHVLLKAKIIGYGVNSTGTNGISNVNLEEQFKERLALYNNNRMDTCV	780
Db	716	KINSAQDLIKNTEHVLLKAKIIGYGVNSTGTNGISNVNLEEQFKERLALYNNNRMDTCV	775
Qy	781	VRNTDDILKACGMAIGDOSMVNPNKYKYLIGKAWNIGISKTANGSKISYVYLGNSPTPE	840
Db	776	VRNTDDILKACGMAIGDOSMVNPNKYKYLIGKAWNIGISKTANGSKISYVYLGNSPTPE	835
Qy	841	NGGNTTNLPNTTNSNARSANALQAPPAQPSATPNLVAI NQHDFGTIESVFELANRSK	900
Db	836	NGGNTTNLPNTTNNARSANYALVNAPPAH-SATPNLVAI NQHDFGTIESVFELANRSK	894
Qy	901	DIDFLYANSQAQGRDLLQTL LIDSHDAGYARKMIDAT SANETTKQLNATPTTLNNIASLE	960
Db	895	DIDFLYHSCQAQGRDLLQTL LIDSHDAGYARQIMDITSTGETTKQLNATDALNNIASLE	954
Qy	961	HKTSGLQTL SLSNAMTILNSRLNLSRRHTNHIDSTAKRLQALKQOKFASLESAAEVLYQF	1020
Db	955	HKTSGLQTL SLSNAMTILNSRLNLSRRHTNNIDSTFARRLQALKQOKFASLESAAEVLYQF	1014
Qy	1021	APKYEKPTNVWANAIGTSLNNGSNASLYGTSGAVDVAIYNGQVEAIYVGGFGSYGSSFNN	1080
Db	1015	APKYEKPTNVWANAIGASLNSGNTSLYGTSGAVDVAIYNEKVEAIYVGGFGSYGSSFNN	1074
Qy	1081	RANSLSNGANTNFVGYSRIFANQHFEDFEAQOALGCSQSSLSNFKSA LLODLNGSYHLA	1140

[illegible]

Db	721	KNSAQDLIKKEHVLLKAKIIGYENASLGTSNISNANLIEOFNERLALYNNNRMDTCV	780
Qy	781	VRNTDDIKACGMAIGDQSMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVLGNSTPTE	840
Db	781	VRNTDDIKACGMAIGDQSMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVLGNSTPTE	840
Qy	841	NGGNTNLTPTNTTSNARSANNAQAONAPPAQSPATPNLVAIINOHDFGTTIESVFELANRSK	900
Db	841	NGGNTNLTPTNTTNARFASALIKNAPFAQTSAIPNLVAINKHDFGTTIESVFELANRSK	900
Qy	901	DIDTLYANGSAGRODLLOTLTLLDSHDAGYARKMIDATSAANEITKQINTATTTLNNTASLE	960
Db	901	DIDTLYANGSAGRODLLOTLTLLDSHDAGYARKMIDATSAANEITKQINTATTTLNNTASLE	960
Qy	961	HKTSGLOTLSLSNAMILNSRLNLSRRHTNHDIDSAKRLQALKDQKFALESAAEVLVYOF	1020
Db	961	HKTSGLOTLSLSNAMILNSRLNLSRRHTNHDIDSAKRLQALKDQKFALESAAEVLVYOF	1020
Qy	1021	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGOVEAIVGGFGYGYSSPNN	1080
Db	1021	APKYEKPTNWNANAIGGASLSNGGNTSLYGTSGVDAYLNGEVEAIVGGFGYGYSSFSN	1080
Db	1081	RANSLNSGANNTFNGVYSRIFANQHEFDEFAOGALGSDOSSLNFKSALLQDLNQSXYLA	1140
Db	1081	QANSLNSGANNANFGVYSRIFANQHEFDEFAOGAGVGSDDSSLNFKDTLLRDLNQSXYLA	1140
Qy	1141	YSAATRASGYDYDFAFRNALVLKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA	1200
Db	1141	YGASARASGYDYDFAFRNALVLKPSGVSYNHLGSTNFKSNSTNKTALKNGASSOHLFNA	1200
Qy	1201	SANVEARYYYGDTSYFYMNAGVLQSFQFAHVGSNNAASLNTFKYNAARNPLNTHARVMGGE	1260
Db	1201	SANVEARYYYGDTSYFYMNAGVLQSFQFANGSSNAVSLNTFKYNAARNPLNTHARVMGGE	1260
Qy	1261	LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF	1296
Db	1261	LQLAKEVFLNLGVVYLHNLISNAGHFNLSNLGMRYSF	1296
RESULT	5		
Q92HV1	PRELIMINARY;	PRT; 1296 AA.	
AC	Q92HV1;		
DT	01-MAY-1999 (TremBLrel. 10, Created)		
DT	01-MAY-1999 (TremBLrel. 10, Last sequence update)		
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)		
DE	VACUOLATING CYTOTOXIN.		
CN	VACA.		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;		
NCBI_TaxID=210;			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=F42;		
RX	MEDLINE=98453456; PubMed=9780260;		
RA	Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,		
RA	Kuriyama M.;		
RT	"Full-length sequence analysis of the vacA gene from cytotoxic and		
RT	noncytotoxic Helicobacter pylori";		
RL	J. Infect. Dis. 178:1391-1398(1998).		
DR	EMBL; AF049626; AAD04265.1;		
DR	InterPro; IPR003842; Vaca.		
DR	Pfam; PF02691; Vaca; 1.		
SQ	SEQUENCE 1296 AA; 139992 MW; AB9E313B29817EFF CRC64;		
Query Match	92.4%;	Score 6165; DB 2; Length 1296;	
Best Local Similarity	91.7%;	Pred. No. 2.8e-228;	
Matches 1189; Conservative	54;	Mismatches 53; Indels 0; Gaps 0;	
Qy	1	MELQOTHRKINRPLVSLALVGLLVSTTPQKSHAAFTTIIIPAIVGGIATGTAVGTSGV	60

Db	1	MELQOTHRKINRPLVSLALVGLLVSTTPQKSHAAFTTIIIPAIVGGIATGTAVGTSGV	60
Qy	61	LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDYLRSLLSKIDGGWDGNAAR	120
Db	61	LGWGLKQAEANKTPDKPKVWRIQAGRCGFNEFPNKEYDYLYKSLLSKIDGGWDGNAAR	120
Qy	121	HYWKGQOQNKLEVDKMDAVGTYTISGLRNFRTGGDLVNMOKATRLGOFNGNSFTSYKD	180
Db	121	HYWKGQOQNKLEVDKMDAVGTYKLSGLRNFRTGGDLVNMOKATRLGOFNGNSFTSFKD	180
Qy	181	SADRTTRVDNFNAKNSIDNFINNVRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD	240
Db	181	SADRTTRVDNFNAKNSIDNFINNVRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD	240
Qy	241	GATNLASSSVKLMGVNMGRLQYVGAYLAPSYSTINTSKVTGVEVFNHILTVGDKNAAQA	300
Db	241	GATNLASSSVKLMGVNMGRLQYVGAYLAPSYSTINTSKVTGVEVFNHILTVGDKNAAQA	300
Qy	301	GIANKKTNIGTLDLWQSAGLNIAPPEGGYKDKPNPTPSQSGAKNDKNSAKNDKQESS	360
Db	301	GIASKKTYIGTLDLWQSAGLNIAPPEGGYKDKPNPTNSQSGAKNDKNSAKNDKQESS	360
Qy	361	QNNSTQVLPNPPNSAQKTEVQPTQVIDGPFAGKDTVVMINRINTNADGTIRVGGFKASL	420
Db	361	QNNSTQVLPNPPNSAQKTEIQPTQVIDGPFAGKDTVVMINRINTNADGTIRVGGYFASL	420
Qy	421	TTNAAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNOVGGYALAGSSANFEE	480
Db	421	TTNAAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGALMVNNOVGGYALAGSSANFEE	480
Qy	481	KAGDTKNGTATFNNDISLGRFVNKLVDAAHTANFKGIDTGNNGGFNTLDFSGVTDKYNIN	540
Db	481	KAGVDTKNGTATFNNNISLGRFVNKLKASHTVNFKIDTGTNGGFFNTLDFSGVTKNYNIN	540
Qy	541	LITASTNVAVKFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLSGGV	600
Db	541	LITASTNVAIKFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSIYSGV	600
Qy	601	KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLTLCQNA	660
Db	601	KFKGGEKLVINDFYYPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLTLCQNA	660
Qy	661	VMDYISOFLNLTIOGDFINNOGTINYLVRGKGVATISVGNAAAMFNNDISATFGYKPLI	720
Db	661	VMDYISOFSNVTIOGDFINNOGTINYLVRGNTIETLSVGNAAVMSFNNDISATFGYKPLI	720
Qy	721	KINSAQDLIKNTEHVLLKAKIIGYENASLGTSNISNANLIEOFNERLALYNNNRMDTCV	780
Db	721	KINSAQDLIKNTEHVLLKAKIIGYENASLGTSNISNANLIEOFNERLALYNNNRMDTCV	780
Qy	781	VRNTDDIKACGMAIGDQSMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVLGNSTPTE	840
Db	781	VRNTDDIKACGMAIGDQAMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVLGNATPAE	840
Qy	841	NGGNTNLTPTNTTSNARSANNAQAONAPPAQSPATPNLVAIINOHDFGTTIESVFELANRSK	900
Db	841	NGGNTNLTPTNTTNARFASALIKNAPFAQTSAIPNLVAINKHDFGTTIESVFELANRSK	900
Qy	901	DIDTLYANGSAGRODLLOTLTLLDSHDAGYARKMIDATSAANEITKQINTATTTLNNTASLE	960
Db	901	DIDTLYANGSAGRODLLOTLTLLDSHDAGYARKMIDATSAANEITKQINTATTTLNNTASLE	960
Qy	961	HKTSGLOTLSLSNAMILNSRLNLSRRHTNHDIDSAKRLQALKDQKFALESAAEVLVYOF	1020
Db	961	HKTSGLOTLSLSNAMILNSRLNLSRRHTNHDIDSAKRLQALKDQKFALESAAEVLVYOF	1020
Qy	1021	APKYEKPTNWNANAIGGTSNLNGSNASLYGTSAGVDAYLNGOVEAIVGGFGYGYSSPNN	1080
Db	1021	APKYEKPTNWNANAIGGASLSNGGNTSLYGTSGVDAYLNGEVEAIVGGFGYGYSSFSN	1080
Qy	1081	RANSLNSGANNTFNGVYSRIFANQHEFDEFAOGALGSDOSSLNFKSALLQDLNQSXYLA	1140
Db	1081	QANSLNSGANNANFGVYSRIFANQHEFDEFAOGAGVGSDDSSLNFKSALLQDLNQSXYLA	1140

OC	Helicobacter.	NCBI_TaxID=210;	Query Match	Score	6162;	DB 2;	Length	1288;
OC	NCBI_TaxID=210;		Best Local Similarity	91.7%;	Pred. No.	3.6e-228;		
RN	[1]		Matches 1189;	Conservative	52;	Mismatches	47;	Indels
RN	SEQUENCE FROM N.A.						8;	Gaps
RP	STRAIN=AFN4769;							
RC	Ji X.H., Rappuoli R., Telford J.L.;							
RA	"Functional analysis of chimeric mutants of the helicobacter pylori							
RT	vaca gene.";							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							
RA	Ji X.H., Rappuoli R., Telford J.L.;							
RT	"Functional analysis of chimeric mutants of the helicobacter pylori							
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF191643; AAK56858.1; -							
SQ	SEQUENCE 1288 AA; 139109 MW; 524ACCF3524FOCA3 CRC64;							
OC	Helicobacter.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=AFN4769;							

```
Qy 121 HYWKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDGLDVMNQKATLRLGQFNGNSFTSYKD 180
Db 121 HYWKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDGLDVMNQKATLRLGQFNGNSFTSFKD 180
Qy 181 SADRTTRVDENAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSCKNAEISLYD 240
Db 181 SADRTTRVDENAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSCKNAEISLYD 240
Qy 241 GATLNLASSSVKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATLNLASNSVKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDRNAQA 300
Qy 301 GIANKKNTNIGTLDLWQASAGLNIIPPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS 360
Db 301 GIATSKKTYIGTLDLWQASAGLNIIPPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS 360
Qy 361 QNNSNTQVIPPNSAQKTEVQPTQIDGPFAGGKDTVVNINRINTNADGIRVGGFKASL 420
Db 361 QNNSSTQVIPPNSGQKTEIQPTQIDGPFAGAKDTVVNINRINTNADGIRVGGYFASL 420
Qy 421 TTNAAHHLHGKGGVNLNQASGRSLIVENLTGNTVDGRLRVNNQVGYALAGSSANFEF 480
Db 421 TTNAANLNICKGGVNLNQASGRSLIVENLTGNTVDGALMVNNQVGYALAGSSANFEF 480
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATNFPGIDTGNNGFNTLDFSGVTDKVNINK 540
Db 481 KAGVDTKNGTATFNNDISLGRFVNLKVDATNFPGIDTGNNGFNTLDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKTNIGISGEVTHFSEDTGSGOSRINTVRLCTRSLSFGV 600
Db 541 LITASTNVAVKFNINELIVKTNIGISGEVTHFSEDTGSGOSRINTVRLCTRSLSFGV 600
Qy 601 KFKGGERLVIDEFYSPWNTFDARNIKVEITNKLFAPGQSPMGTSKLFNNLTQNA 660
Db 601 KFKGGERLVINDFYSPWNTFDARNIKVEITNKLFAPGQSPMGTSKLFNNLTQNA 660
Qy 720 VMDYSQSNLTIGDFTNNGTINLVRGKQVATLSVGNAAAMFNNDISATGYFPLI 720
Db 720 VMDYSQSNLTIGDFTNNGTINLVRGKQVATLSVGNAAAMFNNDISATGYFPLI 720
Qy 780 KINSAQDLINKTEHVLKAKIIGYGVSTGTNGISNVLLEQPKERLALYNNNRMDTCV 780
Db 780 KINSAQDLINKTEHVLKAKIIGYGVSTGTNGISNVLLEQPKERLALYNNNRMDTCV 780
Qy 840 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKNGISKTANGSKISVRYLGNSTPTD 840
Db 840 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKNGISKTANGSKISVRYLGNSTPTD 840
Qy 900 NGGNTTLNPTNTSNARSANALNAQAPPAQSPATPNLVAINHDFGTIESVFELANRSK 900
Db 900 NGGNTTLNPTNTSNARSANALNAQAPPAQSPATPNLVAINHDFGTIESVFELANRSK 900
Qy 960 DIDTLYANSAQGRDLQTLIDSHDAGYARKMIDATANEITKOLNTATTLNNTIASLE 960
Db 901 DIDTLYANSAQGRDLQTLIDSHDAGYARKMIDATANEITKOLNTATTLNNTIASLE 960
Qy 961 HKTSGGLQTLNLSNAMILNRLVNLRRHTNHIDSFARLQALDKQKFALESAAEVLYQF 1020
Db 961 HKTSGGLQTLNLSNAMILNRLVNLRRHTNHIDSFARLQALDKQKFALESAAEVLYQF 1020
Qy 1021 APKYKPTNVWANAIGTSLNNGSNASLYGTSAVDAYLNGOVEAIVGGFGSYGYSFNN 1080
Db 1021 APKYKPTNVWANAIGTSLNNGSNASLYGTSAVDAYLNGOVEAIVGGFGSYGYSFNN 1080
Qy 1081 RANSLNSGANTNPGVYSRIFANQHEFDFAQAGALGSDQSLNFKSALLQDLNQSYYLA 1140
Db 1081 RANSLNSGANTNPGVYSRIFANQHEFDFAQAGALGSDQSLNFKSALLQDLNQSYYLA 1140
Qy 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKNSTNQVALKNGSSQHLFNA 1200
Db 1141 YGASTRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKNSTNQVALKNGSSQHLFNA 1200
Qy 1201 SANVEARYYYGDTSYFYNNAGVQLQEFARHVGSNNAASLNTFKVNAARNPLNTHARYVMGGE 1260
```

```
Db 1201 SANVEARYYYGDTSYFYNNAGVQLQEFANFGSSNAVSLNTEKVNAAARNPLNTHARYVMGGE 1260
Qy 1261 LKLAKEVFLNLGVVYVYLNHLNLSNIGHFASNLGMYRYSF 1296
Db 1261 LKLAKEVFLNLGVVYVYLNHLNLSNIGHFASNLGMYRYSF 1296
RESULT 9
Q92HU8 PRELIMINARY; PRT; 1291 AA.
ID Q92HU8 AC Q92HU8;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=F52;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049631; AAD04269.1; -.
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
DR SEQUENCE 1291 AA; 139473 MW; 398037E7EF290254 CRC64;
Qy 1 MELOOTHRKINRPLVSLALVGLVSIIPQOSHAAFFTVIIPALVGGTATGAVTVSGL 60
Db 1 MELOOTHRKINRPLVSLALVGLVSIIPQOSHAAFFTVIIPALVGGTATGAVTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNEYDLYRSLSSKIDGMDWGNAAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKNEYDLYRSLSSKIDGMDWGNAAAR 120
Qy 121 HYWKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDGLDVMNQKATLRLGQFNGNSFTSYKD 180
Db 121 HYWKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDGLDVMNQKATLRLGQFNGNSFTSFKD 180
Qy 181 SADRTTRVDENAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSCKNAEISLYD 240
Db 181 SADRTTRVDENAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSCKNAEISLYD 240
Qy 241 GATLNLASSSVKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATLNLASNSVKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDRNAQA 300
Qy 301 GIANKKNTNIGTLDLWQASAGLNIIPPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS 360
Db 301 GIATSKKTYIGTLDLWQASAGLNIIPPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS 360
Qy 361 QNNSNTQVIPPNSAQKTEVQPTQIDGPFAGGKDTVVNINRINTNADGIRVGGFKASL 420
Db 359 ---SNTQVIPPNSAQKTEVQPTQIDGPFAGAKDTVVNINRINTNADGIRVGGYFASL 415
Qy 421 TTNAAHHLHGKGGVNLNQASGRSLIVENLTGNTVDGRLRVNNQVGYALAGSSANFEF 480
Db 416 TTNAANLNICKGGVNLNQASGRSLIVENLTGNTVDGALMVNNQVGYALAGSSANFEF 475
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATNFPGIDTGNNGFNTLDFSGVTDKVNINK 540
```

```
Db 476 KAGVDTKNGTIAFNNNISLGRVNLKASHTVNFKIDIDTNGGNTLDFSGVTNKVNINK 535
Qy 541 LITASTNVAVKFNINELIVKNGTISVGEYTHFSDEIGSQSRINTVRLTGTGRSLFSGGV 600
Db 536 LITASTNVAIKFNINELIVKNGTISVGEYTHFSDEIGSQSRINTVRLTGTGRSLFSGGV 595
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPN 660
Db 596 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPN 655
Qy 661 VMDYSQFSNLTTOGDFINNQTINYLVRGKVTATSVGNAAAMFNNDIDATGYKPLI 720
Db 656 VMDYSQFSNLTTOGDFINNQTINYLVRGKVTATSVGNAAAMFNNDIDATGYKPLI 715
Qy 721 KINSQADLLKNTHEVLLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 716 KINSQADLLKNTHEVLLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNY KYLIGKAWKNIGISKTAGSKISVYVIGNSTPTE 840
Db 776 VRNTDDIKACGMAIGDQSMVNNPDNY KYLIGKAWKNIGISKTAGSKISVYVIGNSTPTE 835
Qy 841 NGCNTNTLPTNTTSNARSANNAALQAPPAQSATPNLVAIINQHDFTGTTIESVFELANRSK 900
Db 836 NGCNTNTLPTNTTSNARSANNAALQAPPAQSATPNLVAIINQHDFTGTTIESVFELANRSK 895
Qy 901 DIDTLYANGAGORDLQTLTLLSDHDAGYARKMIDATSAEITKQNTATNTLNNIASLE 960
Db 896 DIDTLYANGAGORDLQTLTLLSDHDAGYARKMIDATSAEITKQNTATNTLNNIASLE 955
Qy 961 HKTSGLOTLSLSNAMILNSRLNRRHTNHIDSFARLQALQKQFASLESAAEVLQF 1020
Db 956 HKTSGLOTLSLSNAMILNSRLNRRHTNHIDSFARLQALQKQFASLESAAEVLQF 1015
Qy 1021 APKYEKPTNVWANAIGTSTLNGSNASLYGTSGAGVDAYLNGOVEALVGGFGSYGSSPNN 1080
Db 1016 APKYEKPTNVWANAIGTSTLNGSNASLYGTSGAGVDAYLNGOVEALVGGFGSYGSSPNN 1075
Qy 1081 RANSLNGANNNTFGYSRIFANQHEFDEAOGALGDSQSSLNFKSALLQDLNQSIVHYLA 1140
Db 1076 QANSLNGANNNTFGYSRIFANQHEFDEAOGALGDSQSSLNFKSALLQDLNQSIVHYLA 1135
Qy 1141 YSAATRASGYDYDFARNALVLPKSPGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1136 YSAATRASGYDYDFARNALVLPKSPGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1195
Qy 1201 SANVEARYYGDTSFYFNAGVLQEFARHVGSNNAASLNTFKYNAARNPLNTHARYVMGGE 1260
Db 1196 SANVEARYYGDTSFYFNAGVLQEFARHVGSNNAASLNTFKYNAARNPLNTHARYVMGGE 1255
Qy 1261 LKLAKEVFLNLGVVYVHLNLSINIGHFASNLGMRYSF 1296
Db 1256 LKLAKEVFLNLGVVYVHLNLSINIGHFASNLGMRYSF 1291
RESULT 10
Q9ZHT8
ID Q9ZHT8 PRELIMINARY; PRT: 1291 AA.
AC Q9ZHT8
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F64;
RX MEDLINE=98453456; PubMed=9780260;
```

```
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049647; AAD04282.1; -.
DR InterPro; IPR003842; VACA.
DR Pfam; PF03691; Vaca; 1.
SQ SEQUENCE 1291 AA; 139395 MW; 4B5785BE5BEE7934 CRC64;
```

Query Match 91.9%; Score 6128.5; DB 2; Length 1291;
Best Local Similarity 91.2%; Pred. No. 6.8e-227;
Matches 1182; Conservative 55; Mismatches 54; Indels 5; Gaps 1;

```
Qy 1 MEIQOQTHRKINRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSG 60
Db 1 MEIQOQTHRKINRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSG 60
Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Qy 121 HYVWKGQOQNKLEVDKMDAVGTYTTLGSLRNFTHGGDLVNMOKATLRLGQFNCGNSFTSYKD 180
Db 121 HYVWKGQOQNKLEVDKMDAVGTYTTLGSLRNFTHGGDLVNMOKATLRLGQFNCGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNTSIDNFEVNNRVGSGAGRKASSTVLTLOASGEGITSDKNAEISLYD 240
Db 181 SADRTTRVDFNAKNTSIDNFEVNNRVGSGAGRKASSTVLTLOASGEGITSDKNAEISLYD 240
Qy 241 GATLNLASSSVKLMGVNMGRLOQYVAYLAPSYSTINTSKVTGEVNFNLTIVGDKNAAQA 300
Db 241 GATLNLASSSVKLMGVNMGRLOQYVAYLAPSYSTINTSKVTGEVNFNLTIVGDKNAAQA 300
Qy 301 GIANKTKTIGTLDLWQSGAGLNIIPAPPEGGYKDKPNPTPSQGAKNKDNESAKNDKQES 360
Db 301 GIANKTKTIGTLDLWQSGAGLNIIPAPPEGGYKDKPNPTPSQGAKNKDNESAKNDKQES 358
Qy 361 QNNSQTVINPPNSAQKTEVQPTVIDGPFAGKDPVAVINRINTNADGTIRVGGFKASL 420
Db 359 ---SNTQVINPPNSAQKTEVQPTVIDGPFAGKDPVAVINRINTNADGTIRVGGFKASL 415
Qy 421 TTNAHLHIKGGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Db 416 TTNAHLHIKGGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 475
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLIKVDATNFKIDTNGGNTLDFSGVTDKVNINK 540
Db 476 KAGVDTKNGTIAFNNNISLGRVNLKASHTVNFKIDIDTNGGNTLDFSGVTDKVNINK 535
Qy 541 LITASTNVAVKFNINELIVKNGTISVGEYTHFSDEIGSQSRINTVRLTGTGRSLFSGGV 600
Db 536 LITASTNVAVKFNINELIVKNGTISVGEYTHFSDEIGSQSRINTVRLTGTGRSLFSGGV 595
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPN 660
Db 596 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPN 655
Qy 661 VMDYSQFSNLTTOGDFINNQTINYLVRGKVTATSVGNAAAMFNNDIDATGYKPLI 720
Db 656 VMDYSQFSNLTTOGDFINNQTINYLVRGKVTATSVGNAAAMFNNDIDATGYKPLI 715
Qy 721 KINSQADLLKNTHEVLLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 716 KINSQADLLKNTHEVLLKAKIIGYGNVSTGNGISNVNLEEQFKERLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNY KYLIGKAWKNIGISKTAGSKISVYVIGNSTPTE 840
Db 776 VRNTDDIKACGMAIGDQSMVNNPDNY KYLIGKAWKNIGISKTAGSKISVYVIGNSTPTE 835
Qy 841 NGCNTNTLPTNTTSNARSANNAALQAPPAQSATPNLVAIINQHDFTGTTIESVFELANRSK 900
Db 836 NGCNTNTLPTNTTSNARSANNAALQAPPAQSATPNLVAIINQHDFTGTTIESVFELANRSK 895
```

```
Db 836 NGGNTTLPNTNATNNARFASVALIKNAPFAQTNPALVAINKHDFGTIESVFELANRSK 895
Qy 901 DIDTLANSAGQGRDLLQTLIDSHDAGYARKMIDATSANETIKOLNTATTTLANNIASLE 960
Db 896 DIDTLAHSGAQGRDLLQTLIDSHDAGYARTMIDATSANETIKOLNTATDALANNIASLE 955
Qy 961 HKTSGLOTLSLSNAMI LNSRLVNLSSRHTNHIDSFARLQALKDQKFALESAAEVLYQF 1020
Db 956 HKTSGLOTLSLSNAMI LNSRLVNLSSRHTNHIDSFARLQALKDQKFALESAAEVLYQF 1015
Qy 1021 APKYEKPTNWANNAIGTSLNNGSNASLYGTSAGVDAYLNCQVEAIVGGFGSYGSFNN 1080
Db 1016 APKYEKPTNWANNAIGTSLNNGSGNTSLYGTSAVDAYLNEKVEAIVGGFGSYGSFNN 1075
Qy 1081 RANSLNGANTNFGVSRIFANOHEFDEFAQAGALGDSQSLNFKSALLDLNOSYHYLA 1140
Db 1076 QANSLNGANNANFGVSRIFANRHEFDEFAQAGAVGDSQSLNFKSALLDLNOSYHYLA 1135
1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1136 YGAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNKTALKNGASSQHLFNA 1195
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFAGVGSNNAAASLNTFKVNAARNPLNTHARVMGMGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFAGVGSNNAAASLNTFKVNAARNPLNTHARVMGMGE 1255
Qy 1261 LKLAKEVFLNLGVYLLHNLISNIGHFASNLGMYSF 1296
Db 1256 LKLAKEVFLNLGVYLLHNLISNIGHFASNLGMYSF 1291

RESULT 11
Q92HT7 ID Q92HT7 PRELIMINARY; PRT; 1296 AA.
AC Q92HT7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
RT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=F68;
MEDLINE=98453456; PubMed=9780260;
Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
Kuriyama M.;
RA "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori."
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049648; AAD04283.1; -.
DR InterPro; IPR003842; VACA.
DR Pfam; PF02691; Vaca; 1.
SQ SEQUENCE 1296 AA; 140049 MW; E66421537383F709 CRC64;
```

```
Query Match 91.88; Score 6122; DB 2; Length 1296;
Best Local Similarity 90.78; Pred. No. 1.2e-226;
Matches 1176; Conservative 65; Mismatches 55; Indels 0; Gaps 0;

Qy 1 MEIOTHRKINRPLVSLALYCALYSITPQOSHAAFFTTVIIPAIVGGIATGTAAGTVSGL 60
Db 1 MEIOTHRKINRPLVSLALYGLLVSIIPQKSHAAFFTTVIIPAIVGGIATGTAAGTVSGL 60

Qy 61 LSWGLKQAEAEANKTPDKPKWRIQAGKGFNEFPKNKYDLYRSLSSKIDGWDGWAAR 120
Db 61 LSWGLKQAEAEANKTPDKPKWRIQAGKGFNEFPKNKYDLYRSLSSKIDGWDGWAAR 120

Qy 121 HYWYKGGQWKNLEVDKDAVGTYKLSGLRNFTGGDLVNMOKATLRLGQFNGNSFTSKD 180
Db 121 HYWYKGGQWKNLEVDKDAVGTYKLSGLRNFTGGDLVNMOKATLRLGQFNGNSFTSKD 180
```

```
Db 121 HYWYKGGQWKNLEVDKDAVGTYKLSGLRNFTGGDLVNMOKATLRLGQFNGNSFTSKD 180
Qy 181 SADRTTTRVDFNAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSBKNAEISLYD 240
Db 181 GTNRTTTRVDFNAKNILIDNFEINNRVSGAGRKASSTVLTLOASEGITSBKNAEISLYD 240
Qy 241 GATLNLASSSYKLMGNVMMGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATLNLASNSYKLMGNVMMGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDRNAQA 300
Qy 301 GIANKKTNITGTLDWOSAGLNIIAPPEGGYKDKPNITPSOSGAKNDKNSAKNDKQESS 360
Db 301 GIATSKTYITGTLDWOSAGLNIIAPPEGGYKDKPNITNSQSGAKSDKNSAKNDKQESS 360
Qy 361 QNNSNTQVINPPNSAQKTEVQPTQIDGPPAGGKDTVVYNNIRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVINPPNSQKTEIQPTQIDGPPAGAKDTVWYNNIRINTNADGTIKVGGYTASL 420
Qy 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTTVDGPLRVNNQVGGYALAGSSANPEF 480
Db 421 TTNAHLNIGKGGVNLNQASGRLLVENLTGNTTVEGALRVNNQVGGYALAGSSANPEF 480
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATANFKGIDTGNCGGFNTLDFSGVTDKVINIK 540
Db 481 KAGVDTKNGTAFNNNISLGRFVNLKASHTVNEFKDIDTGNCGGFNTLDFSGVTNKVINIK 540
Qy 541 LITASTNVAVKFNINELIVKTNIGISVGEYTHFSEIDIGSQSRINTVRLETGTRSLFSGV 600
Db 541 LITASTNVAKFNINELIVKTNIGISVGEYTNFSEIDIGSQSRINTVRLETGTRSIYSGV 600
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKVEITNKLAFGPGQSPWGTSKLPMNLTQLQNA 660
Db 601 KFKGGEKLVINDFYAPWNYFDARNIKVEITNKLAFGPGQSPWGTAKLPMNLTQLPNA 660
Qy 661 VMDYSQFSLTIQGFDFINNOGTINYLVRGGKVATLSVGNAAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLTIQGFDFYNNNOGTINYLVRGGNIETLSVGNAAVMSFNNDIDSATGFYKPLI 720
Qy 721 KINSAODLIKTEHVLLKAKIIGYGNVSTGTNGISNVNLEOFPKRLALYNNNNRMDTCV 780
Db 721 KINSAODLIKKEHVLLKAKIIGYGNVSLGTNSISNANLEOFPNERLALYNNNNRMDTCV 780
Qy 781 VRNTDDIKACGMATIGDSVMNPNPNKYLIKGAWKNTIGISKANGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACGMATIGDQAVNPNPNKYLIKGAWKNTIGISKANGSKISVRYLGNATPPE 840
Qy 841 NGGNTTLPNTTNSARSNANALQAQAPPAQPSATPNLVAINHDFGTIESVFELANRSK 900
Db 841 NGGNTTLPNTTNNARFASVALIKNAPFAQTNPALVAINKHDFGTIESVFELANRSK 900
Qy 901 DIDTLANSAGQGRDLLQTLIDSHDAGYARKMIDATSANETIKOLNTATTTLANNIASLE 960
Db 901 DIDTLAHSGAQGRDLLQTLIDSHNAGYARTMIDATSANETIKOLNANNSALNNIASLE 960
Qy 961 HKTSGLOTLSLSNAMI LNSRLVNLSSRHTNHIDSFARLQALKDQKFALESAAEVLYQF 1020
Db 961 HKTNGLOTLSLSNAMI LNSRLVNLSSRHTNHIDSFARLQALKDQKFALESAAEVLYQF 1020
Qy 1021 APKYEKPTNWANNAIGTSLNNGSNASLYGTSAGVDAYLNCQVEAIVGGFGSYGSFNN 1080
Db 1021 APKYEKPTNWANNAIGTSLNNGSGNTSLYGTSAVDAYLNEKVEAIVGGFGSYGSFNN 1080
Qy 1081 RANSLNGANTNFGVSRIFANOHEFDEFAQAGALGDSQSLNFKSALLDLNOSYHYLA 1140
Db 1081 QANSLNGANNANFGVSRIFANRHEFDEFAQAGAVGDSQSLNFKSALLDLNOSYHYLA 1140
Qy 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1141 YGAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNKTALKNGASSQHLFNA 1200
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFAGVGSNNAAASLNTFKVNAARNPLNTHARVMGMGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVLQEFAGVGSNNAAASLNTFKVNTARNPLNTHARVMGMGE 1260
```

```
QY 1261 LKLAKEVFLNLGVYLHNLISNIGHFASNLGMRYSF 1296
|||||
DB 1261 LKLAKEVFLNLGVYLHNLISNIGHFASNLGMRYSF 1296

RESULT 12
O34108
ID O34108 PRELIMINARY; PRT: 1296 AA.
AC O34108
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE OF 517-760 FROM N.A.
RC STRAIN=F71;
MEDLINE=97339580; PubMed=9196179;
RA Ito Y., Azuma T., Ito S., Miyaji H., Hirai M., Yamazaki Y., Sato F.,
RA Kato T., Kohli Y., Kuriyama M.;
RT "Analysis and typing of the vacA gene from caga-positive strains of
RT Helicobacter pylori isolated in Japan.";
RL J. Clin. Microbiol. 35:1710-1714(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F71;
MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=F71;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF071096; AAC77451.1; -.
SQ SEQUENCE 1296 AA; 139786 MW; F45E2449C811C2BB CRC64;

Query Match 91.8%; Score 6122; DB 2; Length 1296;
Best Local Similarity 90.8%; Pred. No. 1.2e-226;
Matches 1177; Conservative 61; Mismatches 58; Indels 0; Gaps 0;

1 MEIQTHRKINRPLVSLALVGLVSTTPQSHAAFTTIIIPAIYGGIATGTAVGTVSG 60
|||
DB 1 MEIQTHRKINRPLVSLALVGLVSTTPQSHAAFTTIIIPAIYGGIATGTAVGTVSG 60

61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKTDGWDGNGNAAR 120
|||
DB 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKTDGWDGNGNAAR 120

61 LQWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKTDGWDGNGNAAR 120
|||
DB 61 LQWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKTDGWDGNGNAAR 120

121 HYWVGGQONKLEVDKDAVGYTTLISGLRNFTEGGDLVNMQKATLRLGFGNNGSTSYKD 180
|||
DB 121 HYWVGGQONKLEVDKDAVGYTTLISGLRNFTEGGDLVNMQKATLRLGFGNNGSTSYKD 180

121 HYWVGGQONKLEVDKDAVGYTTLISGLRNFTEGGDLVNMQKATLRLGFGNNGSTSYKD 180
|||
DB 121 HYWVGGQONKLEVDKDAVGYTTLISGLRNFTEGGDLVNMQKATLRLGFGNNGSTSYKD 180

181 SADRTTRVDNFNAKNTSIDNFVEINRVSGAGRKASSTVLTILQASEGITSRDAEISLYD 240
|||
DB 181 SADRTTRVDNFNAKNTSIDNFVEINRVSGAGRKASSTVLTILQASEGITSRDAEISLYD 240

241 GATNLASSSVKLMGNVWMLQYVGYLAPSYSTINTSKVTGEVFNHILTVGDKNAQA 300
|||
DB 241 GATNLASSSVKLMGNVWMLQYVGYLAPSYSTINTSKVTGEVFNHILTVGDKNAQA 300

301 GTIANKTNIGTLDLWQSAGLNIAPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
|||
```

```
DB 301 GTIANKTNIGTLDLWQSAGLNIAPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVMINRINTNADGTIRVGGFKASL 420
|||||
DB 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVMINRINTNADGTIRVGGFKASL 420
QY 421 TTNAHLHIGKGGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
|||||
DB 421 TTNAHLHIGKGGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
QY 481 KAGDTKNGTATFNNDISLGRFVNKLKVDATNFKIGDITGNGGFNTLDFSGVTDKVNINK 540
|||||
DB 481 KAGVDTKNGTIAFNNDISLGRFVNKLKVDATNFKIGDITGNGGFNTLDFSGVTDKVNINK 540
QY 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGRSLRSGGV 600
|||||
DB 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGRSLRSGGV 600
QY 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPGQSPWGTSKLMFNLLTGQNA 660
|||||
DB 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPGQSPWGTSKLMFNLLTGQNA 660
QY 661 VMDYSQFSLTQGDPIINQGTINYLVRGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
|||||
DB 661 VMDYSQFSLTQGDPIINQGTINYLVRGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
QY 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGNCISNVNLEEQFKERLALYNNNRMDTCV 780
|||||
DB 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGNCISNVNLEEQFKERLALYNNNRMDTCV 780
QY 781 VRNTDDIKACMAIGDQSMVNNPDNYKYILGKAWKNIGISKTANGSKISVYILNNTPTPE 840
|||||
DB 781 VRNTDDIKACMAIGDQSMVNNPDNYKYILGKAWKNIGISKTANGSKISVYILNNTPTPE 840
QY 841 NGGNTNLTPTNTSNARSANALQAQPAQSPATPNLVAIINQHDPTGTTIESVFELANRSK 900
|||||
DB 841 NGGNTNLTPTNTSNARSANALQAQPAQSPATPNLVAIINQHDPTGTTIESVFELANRSK 900
QY 901 DIDTLYANGSAGORDLLQTLIDSDHAGYARKMIDATSAEITKOLNTATTTLNNTASLE 960
|||||
DB 901 DIDTLYANGSAGORDLLQTLIDSDHAGYARKMIDATSAEITKOLNTATTTLNNTASLE 960
QY 961 HKTSGIQTLSLSNAMILNSRLNRRHTNHIDSFARLQALQDKQFASLESAAEVLQYF 1020
|||||
DB 961 HKTSGIQTLSLSNAMILNSRLNRRHTNHIDSFARLQALQDKQFASLESAAEVLQYF 1020
QY 1021 APKYEKPTNVWANAIGTSTLNGSNSNASLYGTSGAVDAYLNGQVEAIVGGFGSYGSFNN 1080
|||||
DB 1021 APKYEKPTNVWANAIGTSTLNGSNSNASLYGTSGAVDAYLNGQVEAIVGGFGSYGSFNN 1080
QY 1081 RANSLNSGANNTNFGVYSRIFANQHEFDEAOGALGSDOSSLNFKSALLQDLNQSYYLA 1140
|||||
DB 1081 RANSLNSGANNTNFGVYSRIFANQHEFDEAOGALGSDOSSLNFKSALLQDLNQSYYLA 1140
QY 1141 YGASARASYGYDFAFERNALVLPKPSVGNHLSGTNFKSNSTNOVALKNGSSSOHLFNA 1200
|||||
DB 1141 YGASARASYGYDFAFERNALVLPKPSVGNHLSGTNFKSNSTNOVALKNGSSSOHLFNA 1200
QY 1201 SANVEARYYGTSTFYFMNAGVLQEFQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
|||||
DB 1201 SANVEARYYGTSTFYFMNAGVLQEFQFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
QY 1261 LKLAKEVFLNLGVYLHNLISNIGHFASNLGMRYSF 1296
|||||
DB 1261 LKLAKEVFLNLGVYLHNLISNIGHFASNLGMRYSF 1296

RESULT 13
Q9LBC1
ID Q9LBC1 PRELIMINARY; PRT: 1291 AA.
AC Q9LBC1;
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
```

01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN (FRAGMENT).
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHN5060D;
RA Ji X., Telford J.L., Burroni D., Guidotti S., Pagliaccia C.,
RA Rayrat J.M., Xu G., Rappuoli R.;
RT "Allelic variation of vacA gene in the Chinese Helicobacter pylori."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF050328; AAF26510.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
DR NON_TER 1291 1291
SEQUENCE 1291 AA; 139443 MW; 8C525ED343226392 CRC64;

Query Match 91.7%; Score 6120.5; DB 2; Length 1291;
Best Local Similarity 91.1%; Pred. No. 1.4e-226;
Matches 1181; Conservative 56; Mismatches 54; Indels 5; Gaps 1;

QY 1 MEIQOHRKINRPLVSLALVGLVSIITPQSHAAFFTTVIIPAIIVGGIATGTAAGTYSGL 60
DB 1 MELOOHRKINRPLVSLALVGLVSIITPQSHAAFFTTVIIPAIIVGGIATGTAAGTYSGL 60
QY 61 LSWGLKQAEANKTPDKPDKWRIOAQKGFNEFPNKEYDLYRSLLSSKIDGGDWGNAAR 120
DB 61 LGWLKQAEANKTPDKPDKWRIOAQKGFNEFPNKEYDLYRSLLSSKIDGGDWGNAAR 120
QY 121 HYWKGQONKLEVDKMDVGTYLSGLRNFTGGDLVNMOKATLRLQGFNGNSFTSYKD 180
DB 121 HYWKGQONKLEVDKMDVGTYLSGLRNFTGGDLVNMOKATLRLQGFNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKISIDNFVEINNRVSGAGRKASSTVLTQLASEGITSKNAEISLYD 240
DB 181 SADRTTRVDFNAKISIDNFVEINNRVSGAGRKASSTVLTQLASEGITSKNAEISLYD 240
QY 241 GATLNLASSSVKLMGNVWMLQVGLAYLAPSYTINTSKVTGVNFNHLTVGDKNAQA 300
DB 241 GATLNLASSSVKLMGNVWMLQVGLAYLAPSYTINTSKVTGVNFNHLTVGDKNAQA 300
QY 301 GIANKTNTIGTLDLWQSLNIIAPPEGGYKDKPNTTPSQGAKNDKNSAKNDKQESS 360
DB 301 GIASKTYIGTLDLWQSLNIIAPPEGGYKDKPNTTPSQGAKNDKNSAKNDKQD-- 358
QY 361 QNNSNTQVIPPNSAQKTEVQPTQVIGDPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
DB 359 ---SNTQVIPPNSQKTEIQPTQVIGDPFAGGKDTVVNINRINTNADGTIRVGGFKASL 415
QY 421 TTNAAHLHGKGGVNLNQAGRSLLVNTLTGNTVDPGLRVNNOVGYALAGSANPEF 480
DB 416 TTNAAHLHGKGGVNLNQAGRSLLVNTLTGNTVDPGLRVNNOVGYALAGSANPEF 475
QY 481 KAGTDTKNGTATFNNDISLGRFVNLKDAHTANPKGIDTNGGFGNTLDFSGVTDKVNINK 540
DB 476 KAGVDTKNGTATFNNDISLGRFVNLKDAHTANPKGIDTNGGFGNTLDFSGVTDKVNINK 535
QY 541 LITASTVAVKFNINELIVKNTGISVGEYTHFSDIGSQSRINTVLEGTGRSLFSGGV 600
DB 536 LITASTVAVKFNINELIVKNTGISVGEYTHFSDIGSQSRINTVLEGTGRSLFSGGV 595
QY 601 KFKGSEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPMGTSKLMFNNTLQONA 660
DB 596 KFKGSEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPMGTSKLMFNNTLQONA 655
QY 661 VMDYSQFSNLTIOGDFINNOGTINYLVRGGKVATPLSYGNAAMFNNDIDSATGFYKPLI 720
DB 656 VMDYSQFSNLTIOGDFINNOGTINYLVRGGKVATPLSYGNAAMFNNDIDSATGFYKPLI 715

QY 721 KINSAQDLIKNTHVLLKAKIIGYGVNVTGTNGISNVNLEEQFKERLALYNNNNRMDTCV 780
DB 716 KINSAQDLIKNTHVLLKAKIIGYGVNVTGTNGISNVNLEEQFKERLALYNNNNRMDTCV 775
QY 781 VRNTDDIKACGMAIGDOSVWNPNDNYKYLIGKAWNKIGISKANGSKISVYVLCNSTPTE 840
DB 776 VRNTDDIKACGMAIGDOSVWNPNDNYKYLIGKAWNKIGISKANGSKISVYVLCNATPAE 835
QY 841 NGGNTTNLPTNTTSNARSANNAQAAPQSPATPNLVAIINHDFGTIESVFELANRSK 900
DB 836 NGGNTTNLPTNTTNARFARYALIKNAPFACTNATPNLVAIINHDFGTIESVFELANRSK 895
QY 901 DIDTLYANSQAQRDLLOTLIDSHDAGYARKMTIDATSAEITKQLNTATTTLNIIASLE 960
DB 896 DIDTLYANSQAQRDLLOTLIDSHDAGYARTMTIDATSAEITKQLNTATDALNIIASLE 955
QY 961 HKTSGLOTLSISNAMIILNSRLVLSRRHTNHDIFARLQALDKDQFASLESAAEVLYQF 1020
DB 956 HKTSGLOTLSISNAMIILNSRLVLSRRHTNHDIFARLQALDKDQFASLESAAEVLYQF 1015
QY 1021 APYKEKPTNWMANAIGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSGYSSFN 1080
DB 1016 APYKEKPTNWMANAIGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSGYSSFN 1075
QY 1081 RANSLNGANNTPGVYSRIFANQHEPDEFAQAGLSDQSLNFKSALLQDLNOSYHYLA 1140
DB 1076 QANSLNGANNTPGVYSRIFANRHEPDEFAQAGLSDQSLNFKSALLQDLNOSYHYLA 1135
QY 1141 YSARATRSYGVDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
DB 1136 YGATRSYGVDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1195
QY 1201 SANVEARYYYGDTSYFYMNAGVQLQEFARHVSNNAAALSTFKVNAARNPLNTHARVMGGE 1260
DB 1196 SANVEARYYYGDTSYFYMNAGVQLQEFARHVSNNAAALSTFKVNAARNPLNTHARVMGGE 1255
QY 1261 LKLAKEVFLNGVYVNLNLSNIGHFASNLGMRYSF 1296
DB 1256 LQLAKEVFLNGVYVNLNLSNIGHFASNLGMRYSF 1291

RESULT 14
Q9ZHT5 PRELIMINARY; PRT; 1291 AA.
AC Q9ZHT5;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F69;
RA MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RL noncytotoxic Helicobacter pylori."
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049649; AAD04285.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
DR NON_TER 1291 1291
SEQUENCE 1291 AA; 139363 MW; C568C500A2EBE8CF CRC64;

Query Match 91.7%; Score 6117.5; DB 2; Length 1291;
Best Local Similarity 91.2%; Pred. No. 1.8e-226;
Matches 1182; Conservative 56; Mismatches 53; Indels 5; Gaps 1;

Qy 1 MEIQOTHRKNRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSGL 60
Db 1 MELOQOTHRKNRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSGL 60
Qy 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 LGWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Qy 121 HWYKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDLDVNMQKATLRLGQFNGNSFTSYKD 180
Db 121 HWYKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDLDVNMQKATLRLGQFNGNSFTSKD 180
Qy 181 SADRTTRVDFNAKNISIDNFLEINNRVSGAGRKASSTVLTQLQASEGITSGKNAEISLYD 240
Db 181 SADRTTRVDFNAKNISIDNFLEINNRVSGAGRKASSTVLTQLQASEGITSGKNAEISLYD 240
Qy 241 GATNLASSSVKLMGVNMGRLOYVGYAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGVNMGRLOYVGYAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Qy 301 GIANKKTNIGTLDLWQASAGNIAPPEGGYKDKPNTPSQGAKNDKNEKAKNDKQESS 360
Db 301 GIASKKTYIGTLDLWQASAGNIAPPEGGYKDKPNTPSQGAKNDKNEKAKNDKQD-- 358
Qy 361 QNNSTQVINPPNSAQKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 359 ---SNTQVINPPNSAQKTEIQTVIDGPFAGGKDTVVNINRINTNADGTIRVGGYASL 415
Qy 421 TTNAHLHIGKGVNLSNOAGSRSLIVENLTCNITVDGPLRVNNOVGGYALAGSSANFEF 480
Db 416 TTNAANLHIGKGVNLSNOAGSRSLVNLTCNITVEGALMVNNOVGGYALAGSSANFEF 475
Qy 481 KAGDTKNGTATFNNDISILGRFVNKLKDAHTANFKGIDGTGGFNTLDFSGVTDKVNINK 540
Db 476 KAGVDTKNGTATFNNDISILGRFVNKLKASHTVNFKIDGTGGFNTLDFSGVTNKNVINK 535
Qy 541 LITASTNVAVKNFNEILVKTNGTSVGEYTHFSBEDISQSRINTVRLTGTFRSLFSGGV 600
Db 536 LITASTNVAIKNFNEILVKTNGTSVGEYTHFSBEDISQSRINTVRLTGTFRSLFSGGV 595
Qy 601 KFKGGEKLVIDFYYSPNYPDARINKVETNKLAFQPGSPWGTSKLMFNNLTGQNA 660
Db 596 KFKGGEKLVINDFYAPNYPDARINKVETNKLAFQPGSPWGTAKLMFNNLTGQNA 655
Qy 661 VMDYSQFSLNLTQDGFINNQGTINLVRRGGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 656 VMDYSQFSLNLTQDGFINNQGTINLVRRGGKVATLSVGNAAAMFNNDIDSATGYKPLI 715
Qy 721 KINSAQDLINKNKHVLLKAKIIGYENASLGTNSISANLIEQFNERLALYNNNRMDTCV 780
Db 716 KINSAQDLINKNKHVLLKAKIIGYENASLGTNSISANLIEQFNERLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMAIGDQAWNPNPNKYKILGKAWNIGISKTAGSKTSVYVLGNSTPTE 840
Db 776 VRNTDDIKACGMAIGDQAWNPNPNKYKILGKAWNIGISKTAGSKTSVRYLGNATFAE 835
Qy 841 NGGNTNLPTNTTNSARNANLAQNAFPAQSPATPNLVAINQHDGFTIESVFELANRSK 900
Db 836 NGGNTNLPTNTNARFASALIKNAPPAQTNATPNLVAINKHDFGTIESVFELANRSK 895
Qy 901 DIDTLVANSAGQGRDQLLTLLIDSHDAGYARKMIDATSAEITKOLNTATPTLNNIASLE 960
Db 896 DIDTLVANSAGQGRDQLLTLLIDSHDAGYARKMIDATSAEITKOLNTATDALNNIASLE 955
Qy 961 HKTSGLOTLSLNAMILNLSRLVNLSSRRTNTHIDSAKRLQALKDQKFALESAAEVLQF 1020
Db 956 HKTSGLOTLSLNAMILNLSRLVNLSSRRTNTHIDSAKRLQALKDQKFALESAAEVLQF 1015
Qy 1021 APKYKPTNVNANATGCTSLNNGSNALYGTGSAGVDAYLNGQVEAIVGGFSGYGSFNN 1080
Db 1016 APKYKPTNVNANATGCTSLNNGSNALYGTGSAGVDAYLNGQVEAIVGGFSGYGSFNN 1075
Qy 1081 RANSLNSGANNNTFGVYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSXYHIA 1140

Db 1076 QANSLNSGANNNTFGVYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSXYHIA 1135
Qy 1141 YSAATRASGYDFAFERNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSOHLFNA 1200
Db 1136 YGAAATRASGYDFAFERNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSOHLFNA 1195
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSSNAASLNTFKVNAARNPLNTHARVNMGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSSNAASLNTFKVNAARNPLNTHARVNMGE 1255
Qy 1261 LKLAKEVFLNLGVVYVHLNLSINISIGHFASNLGMRYSF 1296
Db 1256 LKLAKEVFLNLGVVYVHLNLSINISIGHFASNLGMRYSF 1291
RESULT 15
O06021 PRELIMINARY; PRT; 1290 AA.
ID O06021;
AC O06021;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE VACUOLATING CYTOFOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43526;
RT "Gura K., Maeda S., Kanai F.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001338; AAB3868.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
SQ SEQUENCE 1290 AA; 139182 MW; 625406E581580578 CRC64;

Query Match 91.6%; Score 6113; DB 2; Length 1290;
Best Local Similarity 91.7%; Pred. No. 2.7e-226;
Matches 1188; Conservative 43; Mismatches 59; Indels 6; Gaps 2;

Qy 1 MEIQOTHRKNRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSGL 60
Db 1 MEIQOTHRKNRPLVSLALVGVLSITPQOASHAAFTTIIIPAIIVGGIATGAVGTVSGL 60
Qy 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Qy 121 HWYKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDLDVNMQKATLRLGQFNGNSFTSYKD 180
Db 121 HWYKGGQONKLEVDMDKAVGTYTTLGSLRNFPGDLDVNMQKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNISIDNFLEINNRVSGAGRKASSTVLTQLQASEGITSGKNAEISLYD 240
Db 181 SADRTTRVDFNAKNISIDNFLEINNRVSGAGRKASSTVLTQLQASEGITSGKNAEISLYD 240
Qy 241 GATNLASSSVKLMGVNMGRLOYVGYAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGVNMGRLOYVGYAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Qy 301 GIANKKTNIGTLDLWQASAGNIAPPEGGYKDKPNTPSQGAKNDKNEKAKNDKQESS 360
Db 301 GIASKKTYIGTLDLWQASAGNIAPPEGGYKDKPNTPSQGAKNDKNEKAKNDKQD-- 355
Qy 361 QNNSTQVINPPNSAQKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 356 QNNSTQVINPPNSAQKTEIQTVIDGPFAGGKDTVVNINRINTNADGTIRVGGYASL 415

QY 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGYALAGSSANPEF 480
Db 416 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGYALAGSSANPEF 475
QY 481 KAGDTKNGTATFNDISLGRFVNLKVDATANFKGIDTGNCGPNTLDFSCVTDKVNINK 540
Db 476 KAGDTKNGTATFNDISLGRFVNLKVDATANFKGIDTGNCGPNTLDFSCVTDKVNINK 535
QY 541 LITASTNVAVKFNINELIYKVTNGISVGEYTHFSEIDIGSQSRINTVRLETGTRSLFSGGV 600
Db 536 LITASTNVAVKFNINELIYKVTNGISVGEYTHFSEIDIGSQSRINTVRLETGTRSLFSGGV 595
QY 601 KFKGGEKLVIDEFYSPWNTFDARNIKNVEITNKLAFCGQSPWGTSKLMFNNTLQONA 660
Db 596 KFKGGEKLVIDEFYSPWNTFDARNIKNVEITNKLAFCGQSPWGTSKLMFNNTLQONA 655
QY 661 VMDYSQFSNLTIOGDFINNOGTINYLVRGKQVATLSVGNAAAMFNNDIDSATGFKPLI 720
Db 656 VMDYSQFSNLTIOGDFINNOGTINYLVRGKQVATLSVGNAAAMFNNDIDSATGFKPLI 715
QY 721 KINSAQDLIKNTEHVLKAKIIGYGVNSTGTNGISNVNLEEQFKERLALYNNNNRMDTCV 780
Db 716 KINSAQDLIKNTEHVLKAKIIGYGVNSTGTNGISNVNLEEQFKERLALYNNNNRMDTCV 775
QY 781 VRNTDDIKACOMAIGDOSMNNPDNYKYLICKAWKNIGISKTANGSKISVYILGNSTPTE 840
Db 776 VRNTDDIKACOMAIGDOSMNNPDNYKYLICKAWKNIGISKTANGSKISVYILGNSTPTE 835
QY 841 NGGNTNLPTNTTSNANNAQAQAPFAOPSATPNLVAINOHDGCTIESVFELANRSK 900
Db 836 NGGNTNLPTNTTSNANNAQAQAPFAOPSATPNLVAINOHDGCTIESVFELANRSK 894
QY 901 DIDPLTHSGAKGRDLLQTLIDSHDAGYVRQMDNTSTGEITKQLNAATTLNNIASLE 960
Db 895 DIDPLTHSGAKGRDLLQTLIDSHDAGYVRQMDNTSTGEITKQLNAATTLNNIASLE 954
QY 961 HKTSGLOTLSLNSAMILNSRLNLSRHTNHIDSFAKRLQALKDQKFALESAAEVLYQF 1020
Db 955 HKTSGLOTLSLNSAMILNSRLNLSRHTNHIDSFAKRLQALKDQKFALESAAEVLYQF 1014
QY 1021 APYKPTNVWANAIGTSLNNGSNASLYGTSGVDAYLNGOVEAIVGGFGSYGYSFNN 1080
Db 1015 APYKPTNVWANAIGTSLNNGSNASLYGTSGVDAYLNGOVEAIVGGFGSYGYSFNN 1074
QY 1081 RANSLSGANTNPGVYSRIFANQHEFDEAQAQALGSDQSSLNPKSALLQDLNQSYHYLA 1140
Db 1075 RANSLSGANTNPGVYSRIFANQHEFDEAQAQALGSDQSSLNPKSALLQDLNQSYHYLA 1134
QY 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1200
Db 1135 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1194
QY 1201 SANVEARYYGDTSYFYMNAGVLQEFAPAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
Db 1195 SANVEARYYGDTSYFYMNAGVLQEFAPAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1254
QY 1261 LKLAKEVFLNLGVVYHLNLSNIGHFASNLGMRYSF 1296
Db 1255 LKLAKEVFLNLGVVYHLNLSNIGHFASNLGMRYSF 1290

Search completed: August 16, 2002, 11:33:59
Job time: 325 sec